NLP
Natural language processing
&lt;/&gt;

Natural Language Processing (NLP) is a branch of computer science, particularly Artificial Intelligence (AI), concerned with enabling computers to interpret and process human language.

The automatic (or semi-automatic) processing of human language is known as natural language processing (NLP).

It is the technology that allows machines to comprehend, analyze, manipulate, and interpret human languages.

The major objective of NLP, in terms of technology, would be to program computers to analyze and interpret massive amounts of natural language data.

Natural Language

A natural language (or ordinary language) is a language that is spoken, written by humans for general-purpose communication.

Example: Hindi, English, French, and Chinese, etc.

A language is a system, a set of symbols, and a set of rules (or grammar).

- The Symbols are combined to convey new information.
- The Rules govern the manipulation of symbols.

Linguistic terminology

Phonetics and phonology	The study of language sounds

Phonetics and phonology	The study of language sounds
Ecology	The study of language conventions for punctuation, text mark-up and encoding

Phonetics and phonology	The study of language sounds
Ecology	The study of language conventions for punctuation, text mark-up and encoding
Morphology	The study of meaningful components of words

Phonetics and phonology	The study of language sounds
Ecology	The study of language conventions for punctuation, text mark-up and encoding
Morphology	The study of meaningful components of words
Syntax	The study of structural relationships among words

Phonetics and phonology	The study of language sounds
Ecology	The study of language conventions for punctuation, text mark-up and encoding
Morphology	The study of meaningful components of words
Syntax	The study of structural relationships among words
Lexical semantics	The study of word meaning

Phonetics and phonology	The study of language sounds
Ecology	The study of language conventions for punctuation, text mark-up and encoding
Morphology	The study of meaningful components of words
Syntax	The study of structural relationships among words
Lexical semantics	The study of word meaning
Compositional semantics	The study of the meaning of sentences

Phonetics and phonology	The study of language sounds
Ecology	The study of language conventions for punctuation, text mark-up and encoding
Morphology	The study of meaningful components of words
Syntax	The study of structural relationships among words
Lexical semantics	The study of word meaning
Compositional semantics	The study of the meaning of sentences
Pragmatics	The study of the use of language to accomplish goals

Phonetics and phonology	The study of language sounds
Ecology	The study of language conventions for punctuation, text mark-up and encoding
Morphology	The study of meaningful components of words
Syntax	The study of structural relationships among words
Lexical semantics	The study of word meaning
Compositional semantics	The study of the meaning of sentences
Pragmatics	The study of the use of language to accomplish goals
Discourse conventions	The study of conventions of dialogue

Formal Language

Before defining formal language Language, we need to define symbols, alphabets, strings, and words.

Symbol is a character, an abstract entity that has no meaning by itself. e.g. Letters, digits, and special characters

Alphabet is a finite set of symbols;

an alphabet is often denoted by **Σ** (sigma)

e.g., **B = {0, 1}** says **B** is an alphabet of two symbols, **0** and **1**.

C = {a, b, c} says **C** is an alphabet of three symbols, **a, b** and c

String or a word is a finite sequence of symbols from an alphabet.

e.g., 01110and 111 are strings from the alphabet B above.

aaabccc and b are strings from the alphabet C above.

Language is a set of strings from an alphabet.

Formal language (or simply language) is a set **L** of strings over some finite alphabet **Σ**.

Formal language is described using formal grammars.

Natural Language Processing (NLP) Components

1. Natural Language Understanding (NLU)
2. Natural Language Generation (NLG)

Natural Language Understanding (NLU):

It extracts metadata from material such as concepts, entities, keywords, emotion, relations, and semantic roles to assist machines in comprehending and analyzing human language.

NLU is mostly used in business applications to comprehend the problem of a client in both spoken and written language.

NLU involves the following tasks -

- It is used to map the given input into useful representation.
- It is used to analyze different aspects of the language.

Natural Language Generation (NLG): It is a translator that translates electronic data into natural language. Text planning, Sentence planning, and Text Realization are the major components.

Steps of Natural Language Processing (NLP)

Lexical Analysis and Morphological: The Lexical Analysis is the initial step in NLP.

The source code is scanned as a stream of characters and converted into meaningful lexemes in this step. The entire text is divided into paragraphs, phrases, and words.

Syntactic Analysis (Parsing): Syntactic analysis is used to examine grammar and word layouts, as well as to show the relationships between words.

Example: Delhi goes to the Rahul

In the real world, Delhi goes to Rahul, which does not make any sense, so this sentence is rejected by the Syntactic analyzer

Semantic Analysis: The representation of meaning is the objective of semantic analysis. The literal meaning of words, phrases, and sentences is the key focus. The semantic analyzer disregards sentences such as “hot ice cream”.

Discourse Integration: It is influenced by the sentences that come before it and evokes the meaning of the words that come after it. For example, the word "that" in the sentence "He wanted that" depends upon the prior discourse context.

Pragmatic Analysis: The fifth and final phase of NLP is pragmatic. It uses a set of principles that describe cooperative talks to assist you in discovering the desired outcome.

Consider the following two sentences:

- The city police refused the demonstrators a permit because they feared violence.
- The city police refused the demonstrators a permit because they advocated revolution.

The meaning of “they” in the 2 sentences is different. In order to figure out the difference, world knowledge in knowledge bases and inference modules should be utilized.

Why NLP is difficult?

NLP is difficult because Ambiguity and Uncertainty exist in the language.

Lexical Ambiguity: It exists in the presence of two or more possible meanings of the sentence within a single word.

Example: Ramya is looking for a match.

In the above example, the word match refers to that either Ramya is looking for a partner or Ramya is looking for a match. (Cricket or other matches)

Syntactic Ambiguity: It exists in the presence of two or more possible meanings within the sentence.

Example: I saw the girl with the binocular.

In the above example, did I have the binoculars? Or did the girl have the binoculars?

Referential Ambiguity: It exists when you are referring to something using the pronoun.

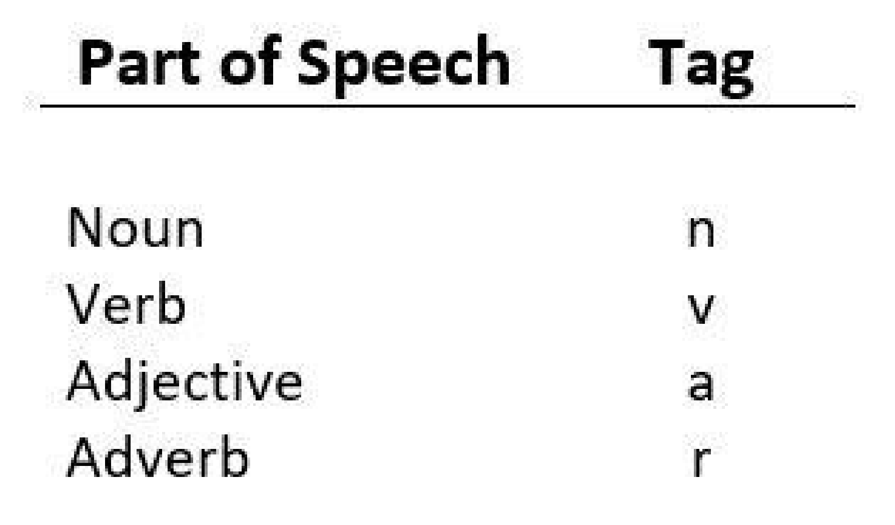
Example: Arun went to Sowmya. She said, "I am hungry."

In the above sentence, you do not know that who is hungry, either Arun or Sowmya.

NLP Techniques

Sentence Splitting: The technique of separating the text into sentences is known as sentence splitting. The technique of dividing the free-flowing text into sentences is known as sentence splitting. It's one of the initial steps in every application that uses natural language processing (NLP). A sentence tokenizer is another name for a sentence splitter. The procedure appears straightforward: after every period, question mark, or exclamation point, put a sentence break.

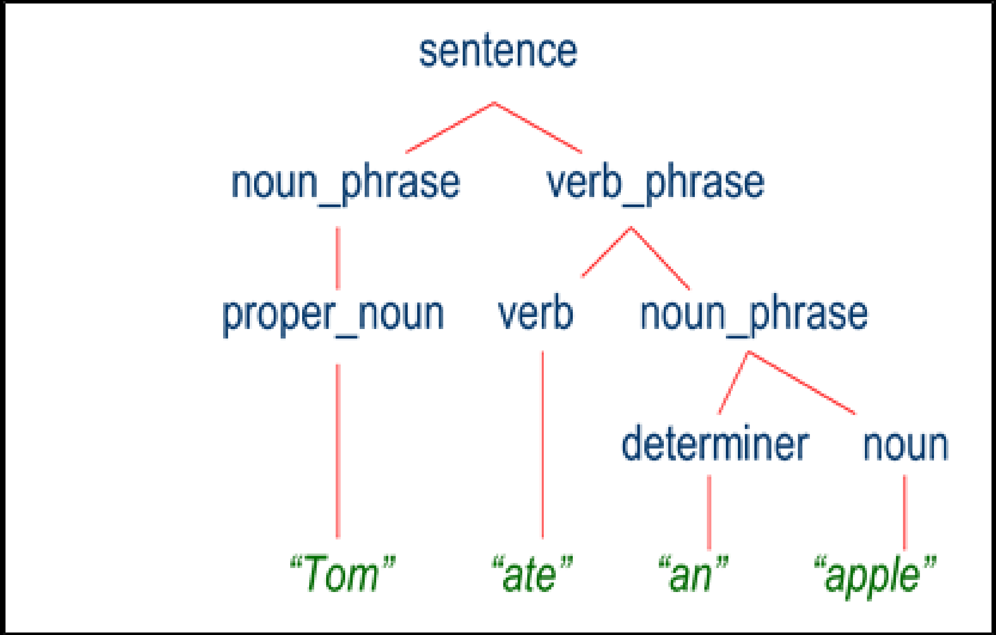
Part-of-speech tagging: It is a method for transforming a phrase into forms, such as a list of words or a list of tuples (each of which has a form (word, tag)). In the case of, the tag is a part-of-speech tag, which indicates whether the word is a noun, adjective, verb, or another type of word.



Part of Speech	Tag
Noun	n
Verb	v
Adjective	a
Adverb	r

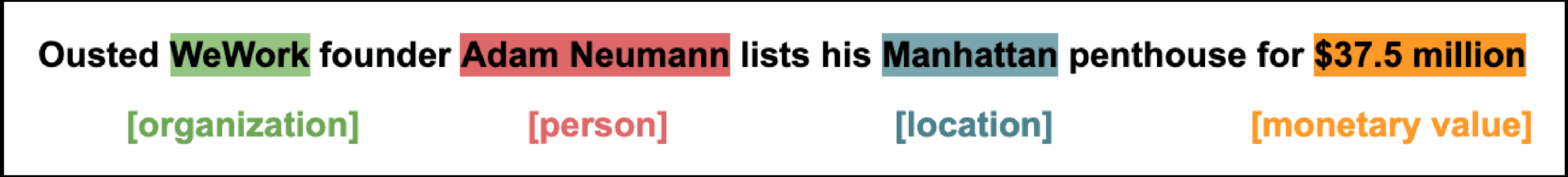
Parsing: In natural language processing, parsing is the process of determining a text's syntactic structure by evaluating its constituent words using an underlying grammar.

Existing parsing methods are mostly based on statistics, probability, and machine learning. The Stanford parser (The Stanford Natural Language Processing Group), and OpenNLP are two noteworthy parsing programs (Apache OpenNLP Developer Documentation)



Named-entity recognition: In any text document, there are particular terms that represent specific entities that are more informative and have a unique context. These are known as named entities, which are words that represent real-world items such as persons, locations, organizations, and so on, and are frequently signified by proper names.

In short, **Named Entity Recognition** is the process of detecting the named entities such as person names, location names, company names, etc from the text. It is also known as entity identification or entity extraction or entity chunking. It is also known as **entity identification** or **entity extraction** or **entity chunking**.



Ousted WeWork founder Adam Neumann lists his Manhattan penthouse for $37.5 million

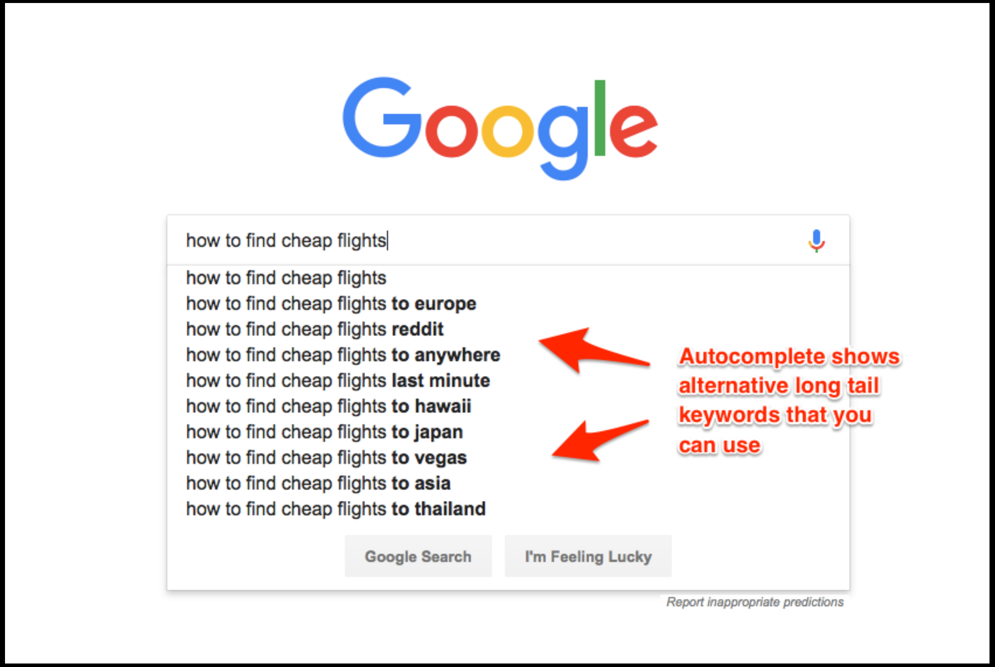
We may extract crucial information to understand the text using named entity recognition, or we can just utilize it to extract important information to record in a database.

Applications of NLP

Voice assistants: A voice assistant is software that uses speech recognition, natural language understanding, and natural language processing to understand the verbal commands of a user and perform actions accordingly.

Example: Siri, Cortana, and Google Assistant.

Auto-complete: Natural Language Processing plays a vital role in grammar checking software and auto-correct functions. Tools like Grammarly, for example, use NLP to help you improve your writing, by detecting grammar, spelling, or sentence structure errors. In search engines (e.g. Google).



how to find cheap flights

how to find cheap flights
how to find cheap flights **to europe**
how to find cheap flights **reddit**
how to find cheap flights **to anywhere**
how to find cheap flights **last minute**
how to find cheap flights **to hawaii**
how to find cheap flights **to japan**
how to find cheap flights **to vegas**
how to find cheap flights **to asia**
how to find cheap flights **to thailand**

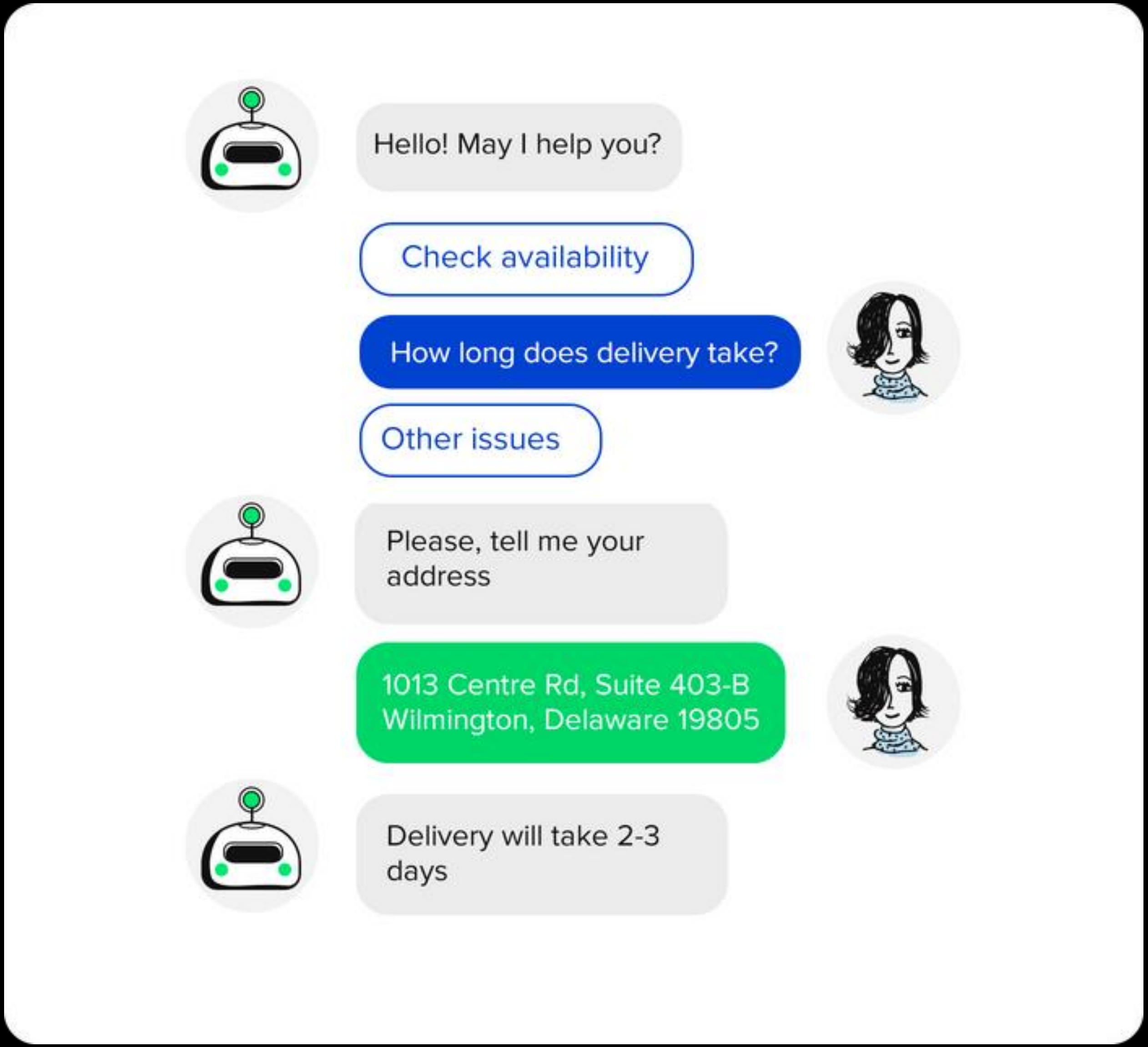
Autocomplete shows alternative long tail keywords that you can use

Spell checking: Almost everywhere, in your browser, your IDE (e.g. Visual Studio), desktop apps (e.g. Microsoft Word).

Machine Translation: Machine Translation is the translation of text or speech by a computer with no human involvement. Machine translation simply performs substitution of words in one language for words in another
Example: Google Translate

Question answering: It focuses on developing systems that can automatically respond to queries asked by humans in their own language. A computer system that understands natural language can use a software system to convert phrases typed by people into an internal representation, allowing the machine to provide legitimate replies. The precise answers can be found by analyzing the questions' syntax and semantics. Some of the problems for NLP in developing excellent question answering systems include lexical gaps, ambiguity, and multilingualism.

Example: Dialog Systems / Chatbots



Hello! May I help you?

How long does delivery take?

Please, tell me your address

1013 Centre Rd, Suite 403-B
Wilmington, Delaware 19805

Delivery will take 2-3 days

Sentiment analysis: is a technique for determining the sentiments of a group of posts. It's also utilized to figure out what people are feeling when they don't say it out loud.

Companies are employing sentiment analysis, a type of natural language processing (NLP), to figure out what their consumers think and feel online. It will assist businesses in determining what their consumers think about their products and services.

With the aid of sentiment analysis, businesses may assess their entire reputation based on consumer posts. In this approach, sentiment analysis goes beyond identifying basic polarity to grasp feelings in context, allowing us to better understand what is underlying the expressed opinion.

NLP Libraries

Natural language Toolkit (NLTK): NLTK is a complete toolkit for all NLP techniques. NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning.

Quepy: Quepy is used to transform natural language questions into queries in a database query language. Quepy is a python framework to transform natural language questions to queries in a database query language. It can be easily customized to different kinds of questions in natural language and database queries.

SpaCy: SpaCy is an open-source NLP library that is used for Data Extraction, Data Analysis, Sentiment Analysis, and Text Summarization.