#### **Tech Accelerator**

A guide to artificial intelligence in the enterprise

<u>Home</u> > <u>Al technologies</u> > <u>Artificial intelligence</u> > natural language processing (NLP)

DEFINITION

# natural language processing (NLP)

Ben Lutkevich, Technical Writer

Ed Burns

### What is natural language processing?

Natural language processing (NLP) is the ability of a computer program to understand human language as it is spoken and written -- referred to as natural language. It is a component of artificial intelligence (AI).

NLP has existed for more than 50 years and has roots in the field of linguistics. It has a variety of real-world applications in a number of fields, including medical research, search engines and business intelligence.

## How does natural language processing work?

NLP enables computers to understand natural language as humans do. Whether the language is spoken or written, natural language processing uses artificial intelligence to take real-world input, process it, and make sense of it in a way a computer can understand. Just as humans have different sensors -- such as ears to hear and eyes to see -- computers have programs to read and microphones to collect audio. And just as humans have a brain to process that input, computers have a program to process their respective inputs. At some point in processing, the input is converted to code that the computer can understand.

There are two main phases to natural language processing: data preprocessing and algorithm development.

Data preprocessing involves preparing and "cleaning" text data for machines to be able to analyze it. preprocessing puts data in workable form and highlights features in the text that an algorithm can work with. There are several ways this can be done, including:

## THIS ARTICLE IS PART OF



Which also includes:

7 kev benefits of AI for business

Criteria for success in Al: Industry best practices

4 main types of artificial intelligence: Explained

- Tokenization. This is when text is broken down into smaller units to work with.
- Stop word removal. This is when common words are removed from text so unique words that offer the most information about the text remain.
- Lemmatization and stemming. This is when words are reduced to their root forms to process.
- Part-of-speech tagging. This is when words are marked based on the part-of speech they are -- such as nouns, verbs and adjectives.

Once the data has been preprocessed, an algorithm is developed to process it. There are many different natural language processing algorithms, but two main types are commonly used:

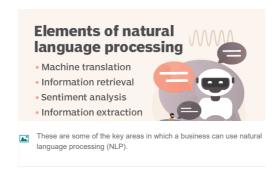
- Rules-based system. This system uses carefully designed linguistic rules. This approach was used early on in the development of natural language processing, and is still used.
- Machine learning-based system. Machine learning algorithms use statistical methods. They learn to perform tasks based on training data they are fed, and adjust their methods as more data is processed. Using a combination of machine learning, deep learning and <a href="neural networks">neural learning</a>. natural language processing algorithms hone their own rules through repeated processing and learning.

## Why is natural language processing important?

Businesses use massive quantities of unstructured, text-heavy data and need a way to efficiently process it. A lot of the information created online and stored in databases is natural human language, and until recently, businesses could not effectively analyze this data. This is where natural

language processing is useful.

The advantage of natural language processing can be seen when considering the following two statements: "Cloud computing insurance should be part of every service-level agreement," and, "A good <u>SLA</u> ensures an easier night's sleep -- even in the cloud." If a user relies on natural language processing for search, the program will recognize that *cloud computing* is an entity, that *cloud* is an abbreviated form of cloud computing and that *SLA* is an industry acronym for service-level agreement.



These are the types of vague elements that frequently appear in human language and that machine learning algorithms have historically been bad at interpreting. Now, with improvements in deep learning and machine learning methods, algorithms can effectively interpret them. These improvements expand the breadth and depth of data that can be analyzed.

#### Techniques and methods of natural language processing

Syntax and semantic analysis are two main techniques used with natural language processing.

Syntax is the arrangement of words in a sentence to make grammatical sense. NLP uses syntax to assess meaning from a language based on grammatical rules. Syntax techniques include:

- Parsing. This is the grammatical analysis of a sentence. Example: A natural language processing algorithm is fed the sentence, "The dog barked." Parsing involves breaking this sentence into parts of speech -- i.e., dog = noun, barked = verb. This is useful for more complex downstream processing tasks.
- Word segmentation. This is the act of taking a string of text and deriving word forms from it. Example: A person scans a handwritten document into a computer. The algorithm would be able to analyze the page and recognize that the words are divided by white spaces.
- Sentence breaking. This places sentence boundaries in large texts. Example: A natural language processing algorithm is fed the text, "The dog barked. I woke up." The algorithm can recognize the period that splits up the sentences using sentence breaking.
- Morphological segmentation. This divides words into smaller parts called morphemes. Example: The word untestably would be broken into [[un[[test]able]]ly], where the algorithm recognizes "un," "test," "able" and "ly" as morphemes. This is especially useful in machine translation and speech recognition.
- Stemming. This divides words with inflection in them to root forms. Example: In the sentence, "The dog barked," the algorithm would be able to recognize the root of the word "barked" is "bark." This would be useful if a user was analyzing a text for all instances of the word bark, as well as all of its conjugations. The algorithm can see that they are essentially the same word even though the letters are different.

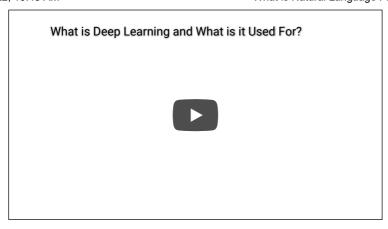
Semantics involves the use of and meaning behind words. Natural language processing applies algorithms to understand the meaning and structure of sentences. Semantics techniques include:

- Word sense disambiguation. This derives the meaning of a word based on context. Example: Consider the sentence, "The pig is in the pen."
  The word pen has different meanings. An algorithm using this method can understand that the use of the word pen here refers to a fenced-in area, not a writing implement.
- Named entity recognition. This determines words that can be categorized into groups. Example: An algorithm using this method could analyze a
  news article and identify all mentions of a certain company or product. Using the semantics of the text, it would be able to differentiate between
  entities that are visually the same. For instance, in the sentence, "Daniel McDonald's son went to McDonald's and ordered a Happy Meal," the
  algorithm could recognize the two instances of "McDonald's" as two separate entities -- one a restaurant and one a person.
- Natural language generation. This uses a database to determine semantics behind words and generate new text. Example: An algorithm could automatically write a summary of findings from a business intelligence platform, mapping certain words and phrases to features of the data in the BI platform. Another example would be automatically generating news articles or tweets based on a certain body of text used for training.

Current approaches to natural language processing are based on deep learning, a type of AI that examines and uses patterns in data to improve a program's understanding. Deep learning models require massive amounts of labeled data for the natural language processing algorithm to train on and identify relevant correlations, and assembling this kind of <u>big data</u> set is one of the main hurdles to natural language processing.

Earlier approaches to natural language processing involved a more rules-based approach, where simpler machine learning <u>algorithms</u> were told what words and phrases to look for in text and given specific responses when those phrases appeared. But deep learning is a more flexible, intuitive approach in which algorithms learn to identify speakers' intent from many examples -- almost like how a child would learn human language.

Three tools used commonly for natural language processing include Natural Language Toolkit (NLTK), Gensim and Intel natural language processing Architect. NLTK is an open source <a href="Python">Python</a> module with data sets and tutorials. Gensim is a Python library for topic modeling and document indexing. Intel NLP Architect is another Python library for deep learning topologies and techniques.



### What is natural language processing used for?

Some of the main functions that natural language processing algorithms perform are:

- Text classification. This involves assigning tags to texts to put them in categories. This can be useful for sentiment analysis, which helps the natural language processing algorithm determine the sentiment, or emotion behind a text. For example, when brand A is mentioned in X number of texts, the algorithm can determine how many of those mentions were positive and how many were negative. It can also be useful for intent detection, which helps predict what the speaker or writer may do based on the text they are producing.
- Text extraction. This involves automatically summarizing text and finding important pieces of data. One example of this is keyword extraction, which pulls the most important words from the text, which can be useful for search engine optimization. Doing this with natural language processing requires some programming it is not completely automated. However, there are plenty of simple keyword extraction tools that automate most of the process the user just has to set parameters within the program. For example, a tool might pull out the most frequently used words in the text. Another example is named entity recognition, which extracts the names of people, places and other entities from text.
- Machine translation. This is the process by which a computer translates text from one language, such as English, to another language, such as French, without human intervention.
- Natural language generation. This involves using natural language processing algorithms to analyze unstructured data and automatically
  produce content based on that data. One example of this is in language models such as <a href="GPT3">GPT3</a>, which are able to analyze an unstructured text and
  then generate believable articles based on the text.



SearchEnterpriseAl



- customer feedback analysis -- where Al analyzes social media reviews;
- customer service automation -- where voice assistants on the other end of a customer service phone line are able to use speech recognition to
  understand what the customer is saying, so that it can direct the call correctly;
- automatic translation -- using tools such as Google Translate, Bing Translator and Translate Me;
- <u>academic research and analysis</u> -- where AI is able to analyze huge amounts of academic material and research papers not just based on the metadata of the text. but the text itself:
- analysis and categorization of medical records -- where AI uses insights to predict, and ideally prevent, disease;
- · word processors used for plagiarism and proofreading -- using tools such as Grammarly and Microsoft Word;
- stock forecasting and insights into financial trading -- using AI to analyze market history and 10-K documents, which contain comprehensive summaries about a company's financial performance;
- · talent recruitment in human resources; and
- automation of routine litigation tasks -- one example is the artificially intelligent attorney.

Research being done on natural language processing revolves around search, especially Enterprise search. This involves having users query data sets in the form of a question that they might pose to another person. The machine interprets the important elements of the human language sentence, which correspond to specific features in a data set, and returns an answer.

NLP can be used to interpret free, unstructured text and make it analyzable. There is a tremendous amount of information stored in free text files, such as patients' medical records. Before <u>deep learning</u>-based NLP models, this information was inaccessible to computer-assisted analysis and could not be analyzed in any systematic way. With NLP analysts can sift through massive amounts of free text to find relevant information.

<u>Sentiment analysis</u> is another primary use case for NLP. Using sentiment analysis, data scientists can assess comments on social media to see how their business's brand is performing, or review notes from customer service teams to identify areas where people want the business to perform better.

## Benefits of natural language processing

The main benefit of NLP is that it improves the way humans and computers communicate with each other. The most direct way to manipulate a computer is through code -- the computer's language. By enabling computers to understand human language, interacting with computers becomes much more intuitive for humans.

### Other benefits include:

- improved accuracy and efficiency of documentation;
- ability to automatically make a readable summary of a larger, more complex original text;
- useful for personal assistants such as Alexa, by enabling it to understand spoken word;

- enables an organization to use chatbots for customer support;
- · easier to perform sentiment analysis; and
- provides advanced insights from analytics that were previously unreachable due to data volume.

## Challenges of natural language processing

There are a number of challenges of natural language processing and most of them boil down to the fact that natural language is ever-evolving and always somewhat ambiguous. They include:

- Precision. Computers traditionally require humans to "speak" to them in a programming language that is precise, unambiguous and highly
  structured -- or through a limited number of clearly enunciated voice commands. Human speech, however, is not always precise; it is often
  ambiguous and the linguistic structure can depend on many complex variables, including slang, regional dialects and social context.
- Tone of voice and inflection. Natural language processing has not yet been perfected. For example, semantic analysis can still be a challenge. Other difficulties include the fact that the abstract use of language is typically tricky for programs to understand. For instance, natural language processing does not pick up sarcasm easily. These topics usually require understanding the words being used and their context in a conversation. As another example, a sentence can change meaning depending on which word or syllable the speaker puts stress on. NLP algorithms may miss the subtle, but important, tone changes in a person's voice when performing speech recognition. The tone and inflection of speech may also vary between different accents, which can be challenging for an algorithm to parse.
- Evolving use of language. Natural language processing is also challenged by the fact that language and the way people use it is continually changing. Although there are rules to language, none are written in stone, and they are subject to change over time. Hard computational rules that work now may become obsolete as the characteristics of real-world language change over time.

## The evolution of natural language processing

NLP draws from a variety of disciplines, including computer science and <u>computational linguistics</u> developments dating back to the mid-20th century. Its evolution included the following major milestones:

- 1950s. Natural language processing has its roots in this decade, when Alan Turing developed the <u>Turing Test</u> to determine whether or not a computer is truly intelligent. The test involves automated interpretation and the generation of natural language as criterion of intelligence.
- 1950s-1990s. NLP was largely rules-based, using handcrafted rules developed by linguists to determine how computers would process language.
- 1990s. The top-down, language-first approach to natural language processing was replaced with a more statistical approach, because
  advancements in computing made this a more efficient way of developing NLP technology. Computers were becoming faster and could be used to
  develop rules based on linguistic statistics without a linguist creating all of the rules. Data-driven natural language processing became mainstream
  during this decade. Natural language processing shifted from a linguist-based approach to an engineer-based approach, drawing on a wider
  variety of scientific disciplines instead of delving into linguistics.
- 2000-2020s. Natural language processing saw dramatic growth in popularity as a term. With advances in computing power, natural language
  processing has also gained numerous real-world applications. Today, approaches to NLP involve a combination of classical linguistics and
  statistical methods.

Natural language processing plays a vital part in technology and the way humans interact with it. It is used in many real-world applications in both the business and consumer spheres, including chatbots, cybersecurity, search engines and big data analytics. Though not without its challenges, NLP is expected to continue to be an important part of both industry and everyday life.

Although there are doubts, natural language processing is making significant strides in the medical imaging field. Learn how radiologists are using Al and NLP in their practice to review their work and compare cases.

This was last updated in March 2021

# **Y** Continue Reading About natural language processing (NLP)

- Al trends in 2020 marked by expectation shift and GPT-3
- GPT3 could revolutionise how business can use Al
- Promethium tool taps natural language processing for analytics
- Chinese Al startup claims breakthrough in pediatric diagnosis
- 5 examples of natural language processing (NLP) in practice

### Related Terms

## conversational Al

Conversational AI is a type of artificial intelligence that enables consumers to interact with computer applications the way they... See complete definition 10

## intelligent system

An intelligent system is an advanced computer system that can gather, analyze and respond to the data it collects from its ... See complete definition 10

## quantum

A quantum (plural: quanta) is the smallest discrete unit of a phenomenon. See complete definition 1

# > Dig Deeper on Al technologies

computational linguistics (CL)

By: Alexander Gillis

speech recognition

By: Ben Lutkevich

natural language generation (NLG)

By: TechTarget Contributor

**Exploring GPT-3 architecture** 

By: George Lawton

-ADS BY GOOGLE

BUSINESS ANALYTICS CIO DATA MANAGEMENT ERP

## **SearchBusinessAnalytics**

## Tableau adds data storytelling tool in latest update

Six months after the analytics vendor's acquisition of Narrative Science comes Data Stories, a new tool aimed at enabling more ...

## New Collibra, Snowflake integration enhances data governance

The new tool aims to enable customers to quickly put protections on sensitive data without requiring them to write code so that ...

About Us Editorial Ethics Policy Meet The Editors Contact Us Advertisers Business Partners Media Kit Corporate Site

Contributors Reprints Answers Definitions E-Products Events Features

Guides Opinions Photo Stories Quizzes Tips Tutorials Videos

All Rights Reserved,
Copyright 2018 - 2022, TechTarget

Privacy Policy

Do Not Sell My Personal Info