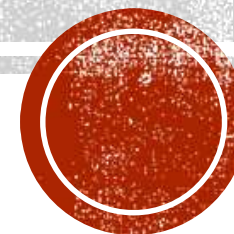


# BASIC NLP



# BASIC NLP

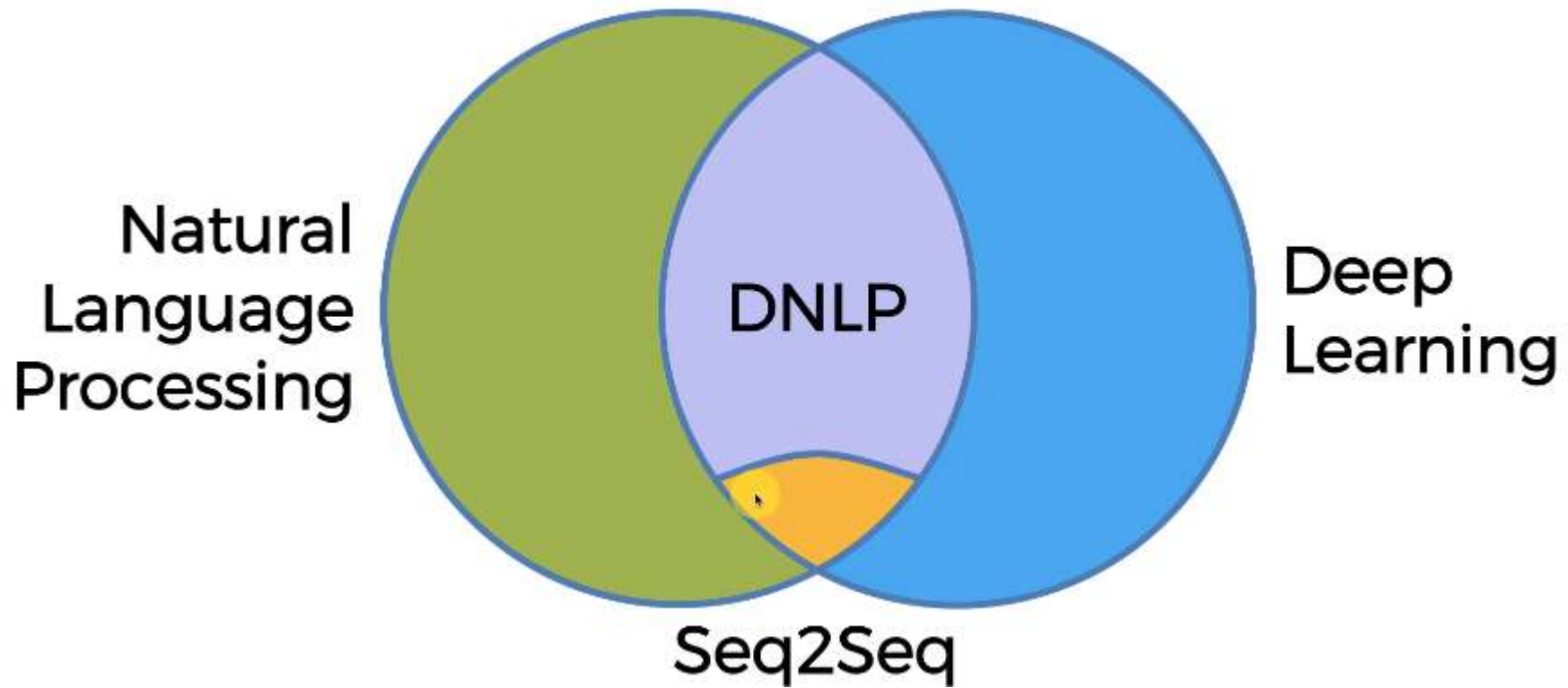
Natural Language Processing (or NLP) is applying Machine Learning models to text and language. Teaching machines to understand what is said in spoken and written word is the focus of Natural Language Processing. Whenever you dictate something into your iPhone / Android device that is then converted to text, that's an NLP algorithm in action.

You can also use NLP on a text review to predict if the review is a good one or a bad one. You can use NLP on an article to predict some categories of the articles you are trying to segment. You can use NLP on a book to predict the genre of the book. And it can go further, you can use NLP to build a machine translator or a speech recognition system, and in that last example you use classification algorithms to classify language. Speaking of classification algorithms, most of NLP algorithms are classification models, and they include Logistic Regression, Naive Bayes, CART which is a model based on decision trees, Maximum Entropy again related to Decision Trees, Hidden Markov Models which are models based on Markov processes.

A very well-known model in NLP is the Bag of Words model. It is a model used to pre-process the texts to classify before fitting the classification algorithms on the observations containing the texts.



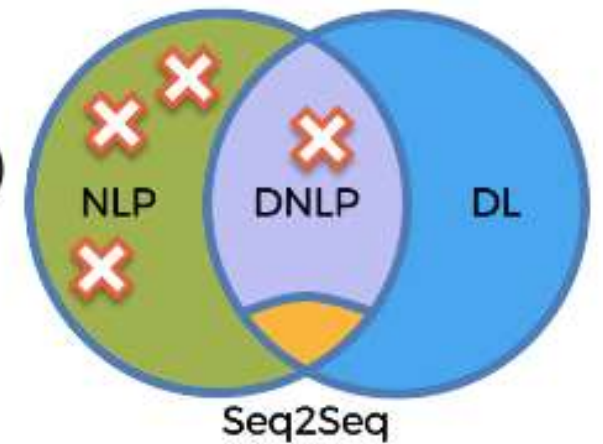
# BASIC NLP



# TYPES OF NLP

Some examples:

1. If / Else Rules (Chatbot)
2. Audio frequency components analysis (Speech Recognition)
3. Bag-of-words model (Classification)
4. CNN for text Recognition (Classification)
5. Seq2Seq (many applications)



# BAG OF WORDS

[0, ..., 0]



20,000 elements long

## 171,476 words

The Second Edition of the 20-volume Oxford English Dictionary contains full entries for **171,476 words** in current use, and **47,158** obsolete words. To this may be added around **9,500** derivative words included as subentries.



How many words are there in the English language?

<https://en.oxforddictionaries.com/.../how-many-words-are-there-in-the-english-language>

[About this result](#) [Feedback](#)

### People also ask

How many words in the English language does the average person know? ^

Most adult native test-takers range from **20,000–35,000 words**. Average native test-takers of age 8 already know **10,000 words**. Average native test-takers of age 4 already know **5,000 words**. Adult native test-takers learn almost 1 new word a day until middle age. May 28, 2013.

Lexical facts - The Economist

<https://www.economist.com/blogs/johnson/2013/05/vocabulary-size>

Search for: How many words in the English language does the average person know?

How many words are there in all the languages in the world? v

Which language has the most words in the world? v

How many words are there in the Spanish language? v

How many words does Eminem say in a minute? v

How many pages is 100 000 words? v



# BAG OF WORDS

Hello Kirill, Checking if you are back to Oz. Let me know if you are around ... Cheers, V

[1, 1, 0, 0, 1, 0, 2, 0, 1, 0, 0, 0, 0, 0, 1, 2, 0, 0, 0, 1, 0, 0, 1, 0, 0, ..., 3]



Yes / No ?



20,000 elements long

## Training Data:

Hey mate, have you read about Hinton's capsule networks?



No

Did you like that recipe I sent you last week?



Yes

Hi Kirill, are you coming to dinner tonight?



Yes

Dear Kirill, would you like to service your car with us again?



No

Are you coming to Australia in December?



Yes

...



...





# BAG OF WORDS

Training Data:

[1, 1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 0, ..., 2]



No

[1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 1, 0, 0, 1, 0, 0, ..., 0]



Yes

[1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, ..., 1]



Yes

[1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, ..., 1]



No

[1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 1, 0, ..., 1]



Yes

...



...

