



Let's Talk NLP: Introduction to NLP

Zephania Reuben

Co-Founder : **TELESOFT AI** & **NILEAGI**

September 3, 2022

Outline

- NLP and It's Components
- NLP Applications in Africa
- General Applications of NLP
- NLP Datasets
- NLP Open Source Tools
- Becoming NLP Engineer
- End Note

NLP and It's Components

Can Computers Understand Language?

What is Natural Language?

- Whatever you speak, read, write, or listen to is mostly in the form of natural language, so it is commonly expressed as natural language.
- For example:
 - The content of this presentation is a source of natural language
 - Movie dialogues are also a source of natural language
 - Your WhatsApp conversations are also considered a form of natural language

What is Natural Language Processing?

- How can you describe the following products.
 - Google Assistant from Google
 - Siri Speech Assistant from Apple
 - Alexa from Microsoft and so on

What is Natural Language Processing?

- Natural Language Processing is a sub-branch of AI.
- Let's say you want to build a machine that interacts with humans in the form of natural language.
- This kind of an intelligent system needs computational technologies and computational linguistics to build it, and the system processes natural language like humans.

What is Natural Language Processing?

- Natural language processing is the ability of computational technologies and/or computational linguistics to process human natural language.
- Natural language processing is a field of computer science, artificial intelligence, and computational linguistics concerned with the interactions between computers and human (natural) languages.
- Natural language processing can be defined as the automatic (or semi- automatic) processing of human natural language.

Components of Natural Language Processing?

- There are two major components of NLP.
 - Natural language understanding
 - Natural language generation

Applications of Natural Language Processing in Africa

- Text Classification
- Text Summarization
- Text Matching/Similarity
- Machine Translation
- Coreference Resolution
- Optical Character Recognition
- Document to Information

General Applications of Natural Language Processing

- Speech recognition system
- Question answering system
- Machine Translation
- Sentiment analysis
- Template-based chatbots
- Topic modeling
- Text Segmentation and Recognition
- Language modeling

Corpus

- Natural language processing related applications are built using a huge amount of data. The large collection of data is called corpus
- Corpus is a collection of written or spoken natural language material, stored on computer, and used to find out how language is used. If you have more than one corpus, it is called corpora.
- In a corpus, the large collection of data can be in the following formats:-
 - Text data, meaning written material
 - Speech data, meaning spoken material

Corpus

- A corpus is also referred to as a dataset in some cases.
- There are three types of corpus:-
 - Monolingual corpus: This type of corpus has one language
 - Bilingual corpus: This type of corpus has two languages
 - Multilingual corpus: This type of corpus has more than one language

Why we need Corpus

- In any NLP application, we need data or corpus to building NLP tools and applications. A corpus is the most critical and basic building block of any NLP-related application.
- Challenges regarding creating a corpus for NLP applications are as follows:-
 - Deciding the type of data we need in order to solve the problem statement
 - Availability of data
 - Quality of the data
 - Adequacy of the data in terms of amount

NLP Open Source Tools

- NLTK: first released in 2001, very broad NLP library
- spaCy: creates parse trees, excellent tokenizer, opinionated
- gensim: topic modeling and similarity detection
- Specialized Tools
 - PyText
 - fastText has library of embeddings
- general ML/DL libraries with text features:
 - sklearn: general purpose Python ML library
 - fastai: fast accurate neural nets using modern best practices, on top of PyTorch
 - keras

How to Become a NLP Engineer/Specialist

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8

