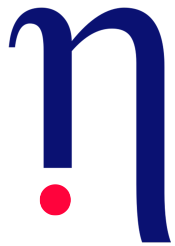
Vema Academy - Zanzibar, Tanzania



Natural Language Processing Ecosystem

Zephania Reuben

Co-Founder : TELESOFT AI & NILE AGI

September 3, 2022

ZINDI

Outline

- NLP and It's Components
- Applications of NLP
- What has been done so far in Africa
- NLP Datasets
- NLP Open Source Tools
- Paths to Become an NLP Engineer
- NLP Engineer Hard Skills
- NLP Engineer Soft Skills
- Responsibilities of a NLP Engineer

ZINDI

Can Computers Understand Human Language?

ZINDI

NLP and It's Components

- 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.

ZiNDi

NLP and It's Components

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

ZiND!

NLP and It's Components

- Natural language processing (NLP) comes from the combination of Computer science, Information sciences, Artificial Intelligence, and Linguistics. NLP focuses on the interaction between computers and human languages.

ZiNDi

NLP and It's Components

- 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.

ZINDI

What is Natural Language Processing?

ZINDI

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.

ZINDI

Components of Natural Language Processing?

- There are two major components of NLP.
  - Natural language understanding(NLU).
  - Natural language generation (NLG).

General Applications of Natural Langauge Processing

- Speech Recognition System.
- Question Answering System.
- Machine Translation.
- Sentiment Analysis.
- Chatbots.
- Topic Modeling.
- Text Summarization.
- Named Entity Recognition.
- Language Modeling.
- Text Classification.
- Text Matching/Similarity.
- Coreference Resolution.
- Optical Character Recognition.

ZiND!

What have been done so far in Africa

- News Classification.
- Machine Translation.
- Sentiment Analysis.
- Speech Recognition.
- Named Entity Recognition.
- Topic Modeling.

Reference here

ZINDI

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: written material.
  - Speech data: spoken material.

ZiNDi

Corpus

- A corpus can also be referred as a dataset.
- 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 languages.

ZINDI

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.

ZiND!

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:
  - Scikit-Learn: general purpose Python ML library.
  - fastai: fast & accurate neural nets using modern best practices, on top of PyTorch.
  - PyTorch.
  - Keras.

ZiND!

Paths to Become an NLP Engineer (Degree vs. Non-degree)

- Many NLP engineers come from an academic background. An associate or bachelor's degree in engineering, data science or computer science is typically preferred. If a bachelor's degree isn't mandatory, a certain amount of experience may be required (including the completion of some NLP courses).
- Earning a certificate is also a good way to gain experience. In some cases, a master's degree or Ph.D. is required.

ZiND!

NLP Engineer Hard Skills

- Understanding of text representation techniques, algorithms, statistics.
- Knowledge of machine learning frameworks and libraries.
- Familiar with Big Data frameworks – Spark, Hadoop.
- Programming skills – Python, Java and/or R.
- Strong problem-solving abilities.
- Syntactic Semantic parsing.
- Knowledge of CI/CD pipelines.
- Strong communication skills.

ZiND!

NLP Engineer Soft Skills

- Creativity.
- Self-motivation.
- Dependability.
- Positivity.
- Conflict management.
- Collaboration.
- Critical thinking.
- Discipline.
- Organization.
- Time management.
- Leadership.

ZiND!

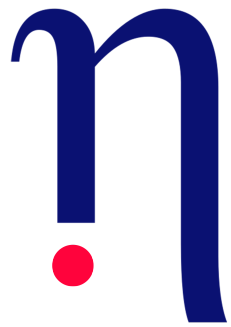
Responsibilities of a NLP Engineer

- Design and develop natural language processing systems.
- Define appropriate datasets for language learning.
- Use effective text representations to transform natural language into useful features.
- Develop NLP systems according to requirements.
- Train the developed model and run evaluation experiments.
- Find and implement the right algorithms and tools for NLP tasks.
- Perform statistical analysis of results and refine models.
- Constantly keep up to date within the field of machine learning.
- Maintain NLP libraries and frameworks.
- Implement changes as needed and analyze bugs.

ZiND!

Thank You!

ZiNDi



ZINDI