

# **A generic NLP (Natural Language Processing) toolset**

The world of Internet is full of free-form, unstructured textual data. This data is an important resource for informed decision making, but its unstructured form creates a challenge.

People realised the need to research and develop ways (algorithms and APIs) to make this textual data machine understandable and make it usable for analysis. That is how quite a few Linguistic APIs came into being, that can be utilised to do just that, namely Stanford, GATE, openNLP etc. Each of these APIs have their pros and cons and most of them are open source (available for free). But, how do you choose the best of what is available out there to serve the purposes of your varied applications?

Well there is a good news for you. Based on industry research of the efficiency and our own experiences working on the Linguistics APIs available in the open source diaspora at the time of taking up this effort, we have assembled a NLP toolset at Synerzip. Our APIs are meant to analyse huge amounts of data and break it into smaller simplified modules which can be recognised easily, by a person or a computer (with preliminary Computational Linguistics related knowledge). The NLP toolset modules accept entire documents as parameters and it returns us meaningful information in a JSON format with extremely high accuracy.

The end users for this NLP toolset would be Big Data analysts, NLP developers, Business Intelligence and Analytics developers.

The NLP toolset provides following APIs related to Natural Language Processing (NLP) or Computational Linguistics.

1. Tokenizer
2. Lemmatizer
3. Sentence Splitter
4. Part of speech (POS) tagging
5. Shallow Parser
6. Deep Parser
7. Named Entity Recognizers
  - A. Gate Annie Named Entity Recognizer
  - B. Apache OpenNLP Named Entity Recognizer
  - C. Stanford Named Entity Recognizer

## **How to Use the NLP Pipeline?**

There are two access points for our API's. The first is built on the Play Framework and the second is a JAR file. However, if you want to have a quick sneak-peek of the functionality, follow [this link](#), enter a sentence or two in the textbox on the left, hit submit and see the results in the box on the right hand side.

### **1. Nlp-Pipeline on Play Framework**

The first form of the APIs is the Nlp-Pipeline on the Play Framework. It provides all the above mentioned APIs in Java on the Play Framework.

For setup and use, please refer to [https://github.com/chetanpalde/nlp\\_pipeline/blob/master/nlp\\_pipeline\\_apis/README.md](https://github.com/chetanpalde/nlp_pipeline/blob/master/nlp_pipeline_apis/README.md)

## **2. Nlp-Pipeline using nlp\_pipeline.jar**

In this scenario, we provide our users the source code which is built in Java and we expect our users to build the nlp\_pipeline.jar file.

For setup, use and build the jar file, please refer to respective fields mentioned in the following link. [https://github.com/chetanpalde/nlp\\_pipeline/blob/master/nlp\\_pipeline\\_source\\_code/README.md](https://github.com/chetanpalde/nlp_pipeline/blob/master/nlp_pipeline_source_code/README.md)

You can use this nlp\_pipeline.jar either by adding it into your project or executing it through the command line.

### **1. Using nlp\_pipeline.jar into your project:**

You can directly import the nlp\_pipeline.jar file in your project and use its methods.

To make this work, you need an NLPApi Object.

### **2. Using nlp\_pipeline.jar through command line:**

You can also run the .jar from a command line. There are two scenarios:

- a. Provide a text file as input
- b. Console based application

For more details, please follow the above mentioned link.

Article By -

Krishna Bhavsar

LinkedIn - <https://in.linkedin.com/in/krishna-bhavsar-31b14914>

Nitin Solanki

LinkedIn - <https://in.linkedin.com/in/nitinmlvya>

Chetan Palde

LinkedIn - <https://in.linkedin.com/in/chetanpalde>