

Parts-of-Speech Tagging – POS Tagger Example in Apache OpenNLP using Java

POS Tagger Example in Apache OpenNLP using Java

POS Tagger Example in Apache OpenNLP marks each word in a sentence with the word type.

An Example :

Input to POS Tagger	John is 27 years old.
Output of POS Tagger	John_NNP is_VBZ 27_CD years_NNS old_JJ ._.

The word types are the tags attached to each word. These Parts Of Speech tags used are from Penn Treebank.

NNP	Proper Noun, Singular
VBZ	Verb, 3rd person singular present
CD	Cardinal Number
NNS	Noun, Plural
JJ	Adjective
.	.

For a complete list of Parts Of Speech tags from Penn Treebank, please refer https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html

Following are the steps to obtain the tags pragmatically in java using apache openNLP

- **Step 1** : Tokenize the given input sentence into tokens.

```
String sentence =  
"John is 27 years old."  
  
String sentence = "John is 27 years old."  
// tokenize the sentence  
tokenModelIn = new FileInputStream("en-token.bin");  
TokenizerModel tokenModel = new TokenizerModel(tokenModelIn);  
Tokenizer tokenizer = new TokenizerME(tokenModel);  
String tokens[] = tokenizer.tokenize(sentence);
```

- **Step 2** : Read the parts-of-speech maxent model, "en-pos-maxent.bin" into a stream.

```
InputStream  
posModelIn = new
```

```
InputStream posModelIn = new FileInputStream("en-pos-maxent.bin");
```

- **Step 3**: Read the stream into parts-of-speech model, POSModel.

```
POSModel posModel =  
new
```

```
POSModel posModel = new POSModel(posModelIn);
```

- **Step 4**: Load the model into parts-of-speech tagger, POSTaggerME

```
POSTaggerME  
posTagger = new
```

```
POSTaggerME posTagger = new POSTaggerME(posModel);
```

- **Step 5**: Grab the tags using the method `POSTaggerME.tag()`, and probability for the tag to be given using the method `PosTaggerME.probs()`;

```
String tags[] =  
posTagger.tag(tokens);
```

```
String tags[] = posTagger.tag(tokens);  
double probs[] = posTagger.probs();
```

- **Step 6**: Finally, print what we got, the token, their respective tags and probabilities of the tags.

The whole program at a glance is given below :

POSTaggerExample.java

```
import  
java.io.FileInputStream  
  
import java.io.FileInputStream;  
import java.io.IOException;  
import java.io.InputStream;  
  
import opennlp.tools.postag.POSModel;  
import opennlp.tools.postag.POSTaggerME;  
import opennlp.tools.tokenize.Tokenizer;  
import opennlp.tools.tokenize.TokenizerME;  
import opennlp.tools.tokenize.TokenizerModel;  
  
/**
```

```
* www.tutorialkart.com
```

```
* POS Tagger Example in Apache OpenNLP using Java
```

```
*/
```

```
public class POSTaggerExample {

    public static void main(String[] args) {

        InputStream tokenModelIn = null;
        InputStream posModelIn = null;

        try {
            String sentence = "John is 27 years old.";
            // tokenize the sentence
            tokenModelIn = new FileInputStream("en-token.bin");
            TokenizerModel tokenModel = new TokenizerModel(tokenModelIn);
            Tokenizer tokenizer = new TokenizerME(tokenModel);
            String tokens[] = tokenizer.tokenize(sentence);

            // Parts-Of-Speech Tagging
            // reading parts-of-speech model to a stream
            posModelIn = new FileInputStream("en-pos-maxent.bin");
            // loading the parts-of-speech model from stream
            POSModel posModel = new POSModel(posModelIn);
            // initializing the parts-of-speech tagger with model
            POSTaggerME posTagger = new POSTaggerME(posModel);
            // Tagger tagging the tokens
            String tags[] = posTagger.tag(tokens);
            // Getting the probabilities of the tags given to the tokens
            double probs[] = posTagger.probs();

            System.out.println("Token\t:\tTag\t:\tProbability\n-----");
            for(int i=0;i<tokens.length;i++){
                System.out.println(tokens[i]+" \t:"\t"+tags[i]+" \t:"\t"+probs[i]);
            }

        }
        catch (IOException e) {
            // Model loading failed, handle the error
            e.printStackTrace();
        }
        finally {
            if (tokenModelIn != null) {
                try {
                    tokenModelIn.close();
                }
                catch (IOException e) {
                }
            }
        }
    }
}
```


- [Named Entity Exatraction](#)
- [Sentence Detection](#)

Learn OpenNLP

- ↳ [OpenNLP Tutorial](#)
- ↳ [Setup Java Project with OpenNLP in Eclipse](#)
- ↳ [OpenNLP Models](#)

Detection / Extraction using Java API

- ↳ [Tokenizer Example](#)
- ↳ [Sentence Detection Example](#)
- ↳ [Parts-Of-Speech Tagger Example](#)
- ↳ [Chunker Example](#)
- ↳ [Lemmatizer Example](#)
- ↳ [Named Entity Extraction Example](#)

Training using Java API

- ↳ [Sentence Detection Model Training](#)
- ↳ [Name Entity Finder Model Training](#)
- ↳ [Document Categorizer Training - Maximum Entropy](#)
- ↳ [Document Categorizer Training - Naive Bayes](#)
- ↳ [Document Categorizer with N-gram features used](#)
- ↳ [Language Detector Training Example](#)

Command Line Tools

- ↳ [Setup and start using Command Line Tools](#)

Useful Resources

- ↳ [How to Learn Programming](#)