

Sentence Detection Example in openNLP using Java

What is Sentence Detection

Sentence Detection or Sentence Segmentation is a process of finding the start and end of a sentence (in a paragraph). This has to be done often in pre-processing section of most of the use cases, which are trying to be solved using Natural Language Processing techniques. Furthermore, Sentence Detection is one of the problems in Natural Language Processing.

Sentence detection is quite challenging because of many reasons in which one of them is : Period symbol (.) which usually denotes the end of a sentence, could also come in an email addresses, abbreviations, decimals etc.,

Sentence Detection Example in openNLP

The following example, SentenceDetectExample.java shows how to use SentenceDetectorME class to detect sentences in a paragraph/string. If you would like to know how to setup eclipse project, refer to [setup of java project with openNLP libraries, in eclipse](#). The process should be same, even for a different IDE(adding the required jars to the build path should do the magic).

SentenceDetectExample.java

```
import
java.io.FileInputStream

import java.io.FileInputStream;
import java.io.IOException;
import java.io.InputStream;

import com.fasterxml.jackson.databind.exc.InvalidFormatException;

import opennlp.tools.sentdetect.SentenceDetectorME;
import opennlp.tools.sentdetect.SentenceModel;

/**
 * Sentence Detection Example in openNLP using Java
 * @author tutorialkart
 */
public class SentenceDetectExample {

    public static void main(String[] args) {
        try {
            new SentenceDetectExample().sentenceDetect();
        } catch (IOException e) {
            e.printStackTrace();
        }
    }
}
```

```

}

/**
 * This method is used to detect sentences in a paragraph/string
 * @throws InvalidFormatException
 * @throws IOException
 */
public void sentenceDetect() throws InvalidFormatException, IOException {
    String paragraph = "This is a statement. This is another statement. Now is an abstract word for time, that is always
flying.";

    // refer to model file "en-sent.bin", available at link http://opennlp.sourceforge.net/models-1.5/
    InputStream is = new FileInputStream("en-sent.bin");
    SentenceModel model = new SentenceModel(is);

    // feed the model to SentenceDetectorME class
    SentenceDetectorME sdetector = new SentenceDetectorME(model);

    // detect sentences in the paragraph
    String sentences[] = sdetector.sentDetect(paragraph);

    // print the sentences detected, to console
    for(int i=0;i<sentences.length;i++){
        System.out.println(sentences[i]);
    }
    is.close();
}
}

```

When SentenceDetectExample.java is run, the console output is :

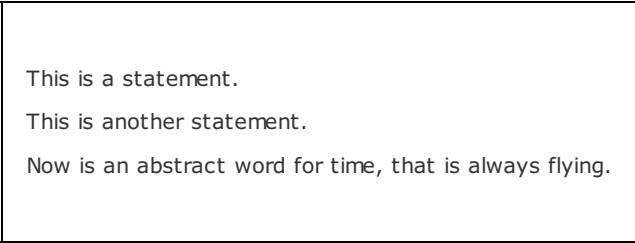
Program Output



```

This is a statement.
This is another

```

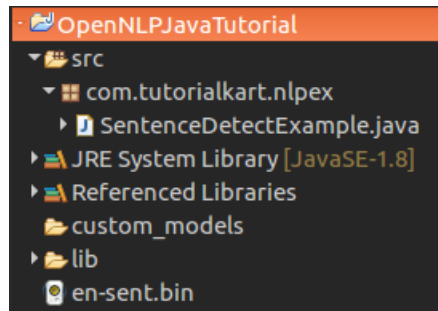


```

This is a statement.
This is another statement.
Now is an abstract word for time, that is always flying.

```

The project structure and model file location, etc., for the example is shown below:



Example Project – Structure

Model File:

The model file en-sent.bin is available at [<http://opennlp.sourceforge.net/models-1.5/>]. Stay updated regarding latest releases of openNLP or model files, at [<https://opennlp.apache.org/download.html>]

Java Documentation

Find the java documentation for SentenceDetectorME at official site and play with the other methods like `getSentenceProbabilities()`, `sentPosDetect(String s)`, etc., for a better understanding.

Custom model for Sentence Detection from user defined training data

If you are interested in knowing of how to train and generate a model yourself for Sentence Detection, refer to [training a model for Sentence Detection in openNLP](#).

Conclusion :

In this openNLP tutorial, we have seen Sentence Detection Example in openNLP using Java.

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](#)