

# Freeling interface

## Description of the interface

The main objective of this interface is to tokenize a given text, and to obtain a sort of properties for each word. The interface is developed under the file “freeling\_functions.py”.

## The approach

In order to use Freeling, this interface considered the use of the sample program called “analyze”. It is a sample program of Freeling that given a sort of parameters can return an output with the desired result. In order to keep the result simple but useful and prevent extra installation files, it is used the default configuration file available after the clean installation for every language supported. For example the configuration file for the english language is available under the folder “/usr/local/share/freeling/en.cfg”.

The use of the program “analyze” against the use of an own code for analyzing the text has been chosen in order to keep the simplicity of the program, and because it is already provided with the installation.

## Using the interface “freeling\_functions.py”

This interface is composed by 5 functions that return lists of words and its characteristics.

### **tokenize\_file(path\_to\_file, lang='en')**

The output of the sentence “I'm studying a master in Barcelona.” would be a list like the following (word, lemma, category, probability assignment):

```
[[['I', 'i', 'PRP', '1'], ['am', 'be', 'VBP', '1'], ['studying', 'study', 'VBG', '1'], ['a', '1', 'Z', '0.99998'],  
['master', 'master', 'NN', '0.909774'], ['in', 'in', 'IN', '0.985534'], ['Barcelona', 'barcelona', 'NP', '1'],  
['.', '.', 'Fp', '1']]
```

### **extract\_type(path\_to\_file, lang='en')**

Returns a list of words with the corresponding category of word. i.e. 'NP', 'VBG',...

### **extract\_phonetics(path\_to\_file, lang='en')**

Returns the word and its corresponding phonetics notation.

### **extract\_probabilities(path\_to\_file, lang='en')**

Returns the word and the probability of assignment of each word.

### **get\_names(path\_to\_file, lang='en')**

Returns a list of the words considered as a name 'NP'.

## **Evaluation of the results**

Freeling is a powerful tool that can provide many options, but as the main objective was to provide a simple reference of each word, the approach used can bring a simple solution to the developer that needs to use it.

## **Proposal of improvements**

As an interface, the main goal is to provide information about the given text, and using this interface this can be achieved. If extra information is needed in the future, it can be considered as an improvement, adding some extra functions to provide further information about a sentence, or a given word. An example of improvement could be to let the user define his own configuration file and define the properties that want to be shown.