

KEX3 Documentation

The program contains two classes: KEX3 and Extractor

KEX3 has the methods that are the building blocks of our models and the Extractor Class is like a wrapper class that simplifies this process.

The Extractor Class has six modules:

- 1) Clustering with number of keywords relative to the input document
- 2) Skip-Agglomeration with the number of keywords relative to the input document
- 3) Clustering with an absolute number of keywords
- 4) Skip-Agglomeration with an absolute number of keywords
- 5) Clustering with an absolute number of keywords along with an initial word (Skips the selection part)
- 6) Skip-Agglomeration with an absolute number of keywords along with an initial word (Skips the selection part)

The Extractor class is imported as follows:

```
from KEX3 import Extractor
```

The Extractor Class object takes three arguments: the stop-word list, word vectors, pre-trained TF-IDF scores dictionary.

Creating the Class Object

```
ext = Extractor(stopwords, word_vectors, tf_idf_scores)
```

Accessing the methods using this Object

The Extractor class has six unique methods that are representatives of the six modules shown above. The execution procedure for each of these methods is given below.

Clustering with number of keywords relative to the input document

```
ext.clustering_keywords(text, metric)
```

text – the input text document

metric – one of the three selection techniques (avg – average diameter, chain – chain length, skip – skip-agglomerative distance)

Skip-Agglomeration with number of keywords relative to the input document

ext.skip_agglomeration_keywords(text, metric)

text – the input text document

metric – one of the three selection techniques (avg – average diameter, chain – chain length, skip – skip-agglomerative distance)

Clustering with an absolute number of keywords

ext.clustering_keyphrases(text, metric, number of keywords)

text – the input text document

metric – one of the three selection techniques (avg – average diameter, chain – chain length, skip – skip-agglomerative distance)

number of keywords – number of keywords/key-phrases to return irrespective of the document size.

Skip-Agglomeration with an absolute number of keywords

ext.skip_agglomeration_keyphrases(text, metric, number of keywords)

text – the input text document

metric – one of the three selection techniques (avg – average diameter, chain – chain length, skip – skip-agglomerative distance)

number of keywords – number of keywords/key-phrases to return irrespective of the document size.

Clustering with an absolute number of keywords along with an initial word (Skips the selection part)

ext.clustering_with_related_word(text, number of keywords, initial/related word)

text – the input text document

number of keywords – number of keywords/key-phrases to return irrespective of the document size.

Initial/related word – selects the cluster from a set of clusters whose center is the given word.

Skip-agglomeration with an absolute number of keywords along with an initial word (Skips the selection part)

```
ext.skip_agglomeration_with_related_word(text, number of  
keywords, initial/related word)
```

text – the input text document

number of keywords – number of keywords/key-phrases to return irrespective of the document size.

Initial/related word – selects the cluster from a set of clusters whose center is the given word.