

Content

1	General.....	1
2	Installation.....	1
3	Quick Guide.....	2
4	Windows.....	2
4.1	Main.....	2
4.2	Main table of main window.....	3
4.3	Word Count.....	3
4.4	Sentence Display.....	4
4.5	Settings.....	5
4.6	Credits.....	6

1 General

Analyze the content of a text. Diagonal reading with subject, predicate and object. Search for similar pdf or plain text content using Bing search.

Supports English, French and German text and pdf files.

Portable application. Runs under Windows and Linux. JavaFX application.

Open source: [Apache License Version 2.0](#)

Part of speech tagging builds upon Apache OpenNLP.

NLP (Natural Language Processing) extensions:

- Divide sentences in sub clauses: segmentation.
- Divide plain text: subject, predicate, object.
- Count words: stemming.
- Search for similar content: pdf's.

[Download](#) [Wiki](#) [Javadoc](#) [GitHub](#)

Brought to you by

open-hopin

Buelow Sensalgo

2 Installation

Unzip to destination folder. Open mbFXWords.exe for Windows or mbFXWords.sh for Linux.

Make sure JRE 1.8 or higher is installed on your machine. (Java Runtime Environment)

Command line optionally accepts file to be analyzed with relative or absolute

path.

Example files to be analyzed can be found in sub directory “resources”.

mbFXWords can be used for pdf displayed in an Internet browser. For Firefox integration install mbFXWords to Firefox installation directory. Firefox must be able to display pdf on a tab. Then download and open with mbFXWords from there.

3 Quick Guide

For a first impression simply click Analyze (text “resources/default.txt”). Choose a different file by button Choose. A selection of content clauses (subject, predicate and object) you find in the middle of main window, a content’s abstract at the top and most common words at the bottom.

4 Windows

4.1 Main

The screenshot shows the 'mbFXWords - analyze a text - diagonal read pdf's' window. It features a text input area at the top with a 'Choose pdf/txt' button and a 'C2C' button. Below the text is a table with columns for 'subject', 'predicate', 'object', and 'sentence'. The table contains several rows of text analysis results. On the left side, there are buttons for 'Analyze', 'Search for similar', 'Settings', and 'About + Credits'. At the bottom, there is a search bar and a page indicator showing '1'.

Callouts in the image include:

- abstract**: points to the text input area.
- copy abstract to clipboard**: points to the C2C button.
- open different file**: points to the 'Choose pdf/txt' button.
- manual**: points to a '?' button.
- see explanation under 'word count'**: points to the 'number of different stems (words) effective' section.
- total number of clauses in the text**: points to the 'clauses 19' label.
- part of subjects, predicates and objects known by program**: points to the table.
- Bing search in browser**: points to the 'Search for similar' button.
- main table**: points to the main table area.
- turn pages of main table**: points to the page indicator.
- most common words in respective order**: points to the search bar.

4.2 Main table of main window

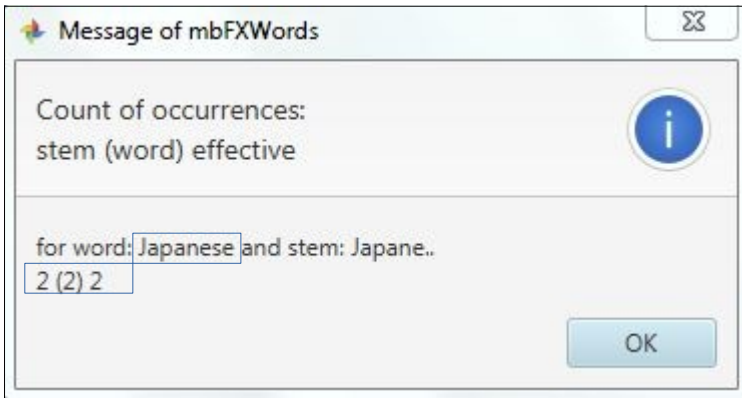
Division of complete text in sub clauses.

words	subject	predicate	object	sentence
12	Japanese	have	punctuality	1
2	maintenance	separated	freight	2E
1	Deutsche	cut	slice	4
		dream	trains	7S
		is	way	8
12	rail	induced	passengers	9
11	Tsukuba	drove	9:44	13
15	customer	said	website	17

Callouts in the image include:

- 12 words form the first clause**: points to the first row of the table.
- sub clause no. 4, consecutive numbering, link to complete sentence**: points to the '4' in the 'sentence' column.
- predicate, link to word count**: points to the 'is' in the 'predicate' column.

4.3 Word Count

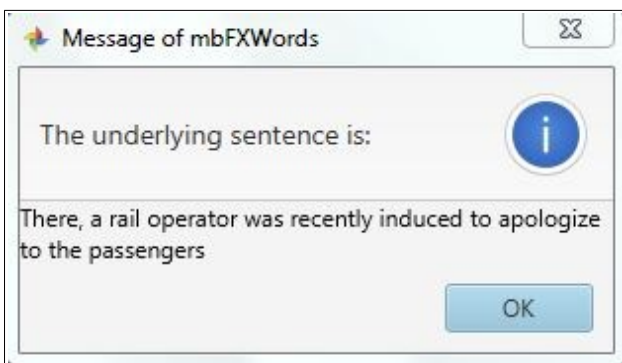


The word “Japanese” in this inflected form occurs two times in complete text. This is the number in braces.

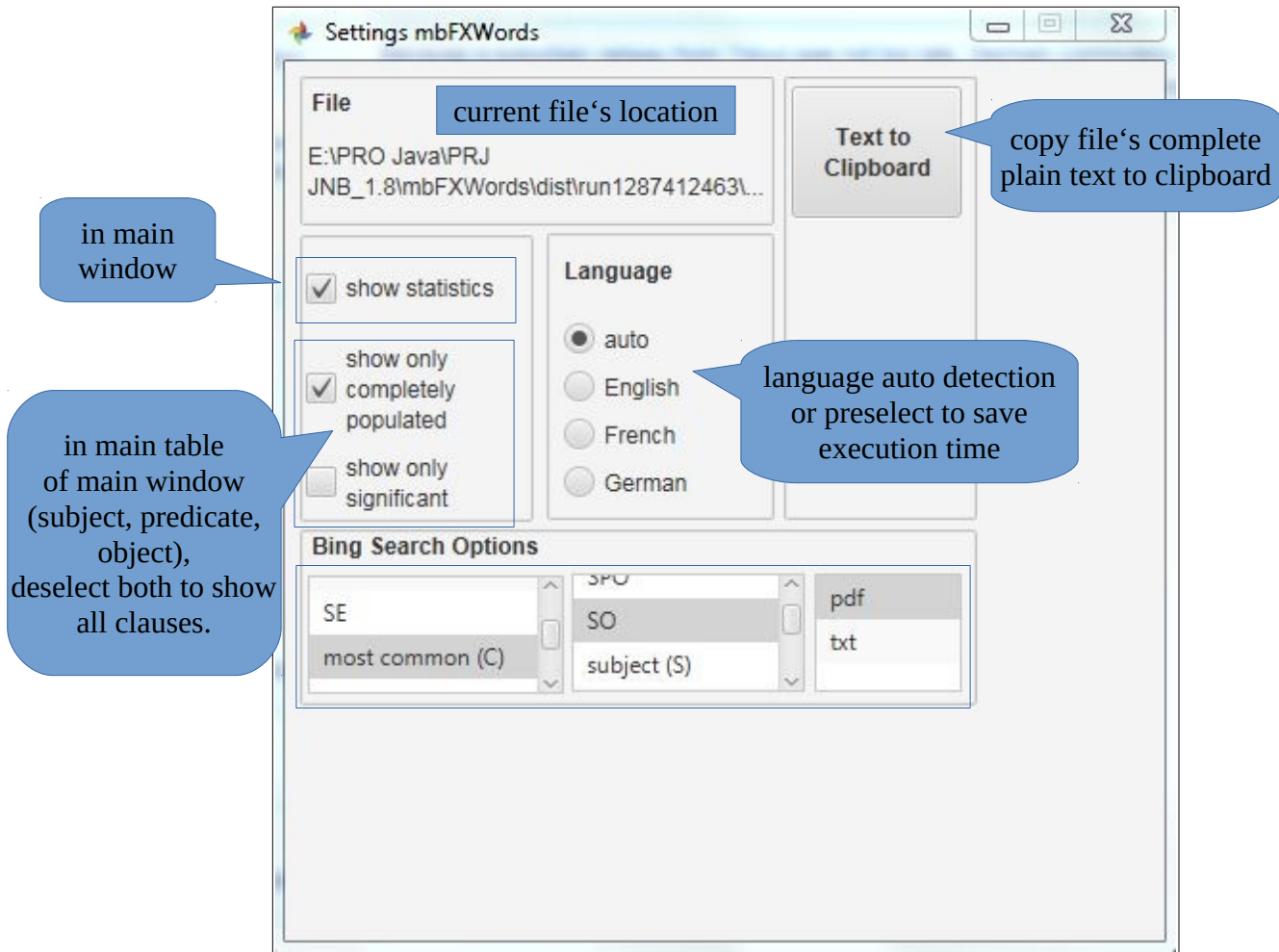
The stem is the word with last two letters substituted by points, so words that differ in last two letters have the same stem. In complete text stem occurs 2 times. This is the first number.

If words can differ in length by two letters and it does not matter which letter a point in the stem stands for, we have the number of occurrences of such words as indicated by last number.

4.4 Sentence Display



4.5 Settings



Search in browser for pdf or txt (select on the right) file with similar content to loaded file according to following options:

(select on the left)

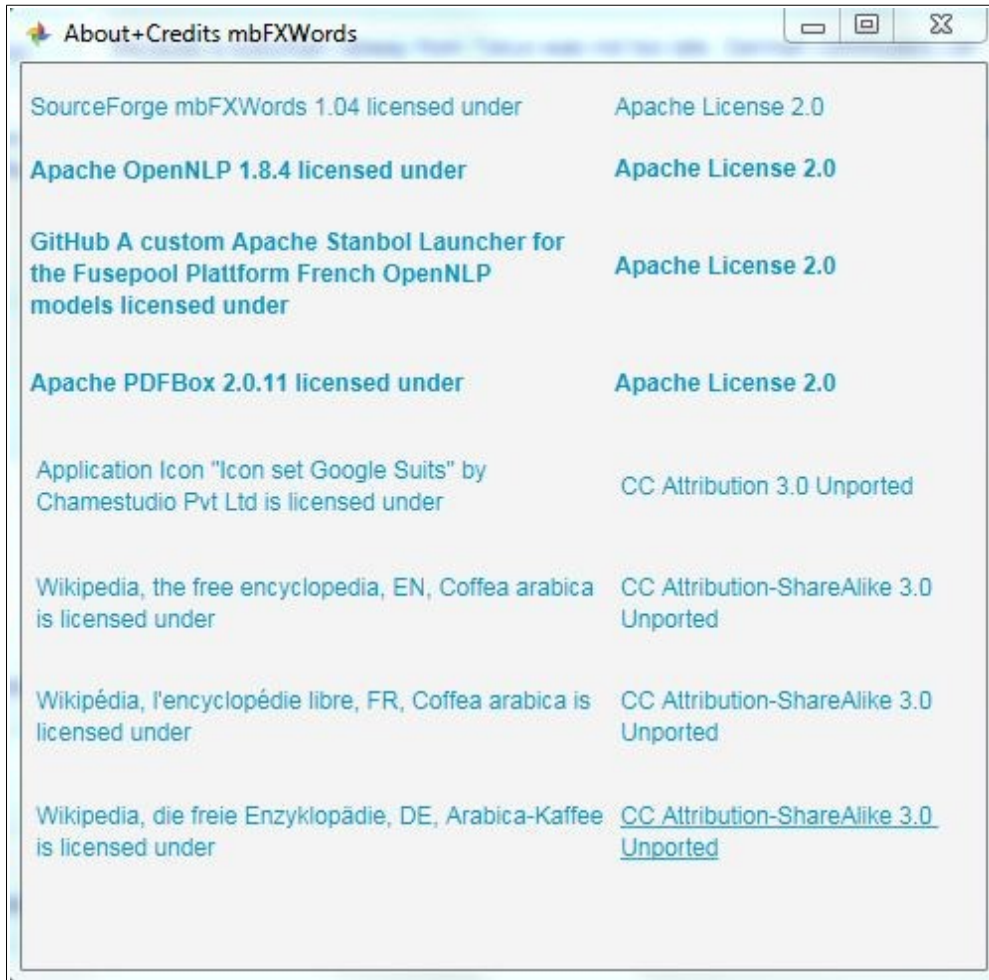
CSE – common, significant and extraordinary
 CS – common and significant
 SE – significant and extraordinary
 most common (C) – only this
 significant (S) – only this
 extraordinary (E) – only this

(select in the middle)

SPO – search for subject, predicate and object
 SO – search for subject and object
 subject (S) – only this
 predicate (P) – only this
 object (O) – only this

All columns form a combined search string, for example for the predicate that is most common in a pdf file.

4.6 Credits



Used modules are part of main application and its source libs. No extra installation necessary.