27/06/2018

User guide

UCSI software project

University of Colorado Summer Internship (UCSI) software project user guide

Table des matières

Introduction

This app is a visualization tool for Natural Language Processing. It has two purposes:

- Process textual data and analyze it
- Visualize the data obtained after analysis

The app should help researchers in linguistics understanding the structure of the documents they study. It provides a set of statistical metrics, sentence structure analysis and other analysis. It then represents them in order to extract useful information from it. It is built on Shneiderman's mantra "Overview first, zoom and filter, then details-on-demand". It should cover a wide range of the possible analysis in Natural Language Processing and present them efficiently.

I.   Installation

a. R version

In order to use the application, you need to have the 3.5.0 (2018-04-23) R version, which is currently the latest R version. It should work with updates, so if you have a later version, it should also run.

b. Packages to install

In order to use the application, you also need specific packages. A R program will be developed to facilitate the installation of the software. However, currently you need to install the following packages:

- *shiny*
- *DT*
- *plotly*
- *wordcloud2*
- *tm*
- *crosstalk*
- *webshot*
- *phantomjs*
- *tinytex*
- *shinydashboard*

The packages needed will expand as the back-end evolves. Indeed, the application doesn't really process the data yet (only the minimum to test the app), so there will be new packages to install in order to perform this task.

Firstly, you need to install Rtools, which can be done at https://cran.r-project.org/bin/windows/Rtools/ .

To install the packages, you need to copy/paste the following lines into the console:

```
install.packages("shiny")
install.packages("DT")
install.packages("plotly")
install.packages("wordcloud2")
install.packages("tm")
install.packages("crosstalk")
install.packages("webshot")
```

```r
webshot::install_phantomjs()
install.packages("tinytex")
tinytex::install_tinytex()
install.packages("wordcloud")
install.packages("rmarkdown")
install.packages("knitr")
install.packages("httpuv")
install.packages("stringi")
install.packages("shinydashboard")
```

If that fails, you can try to install also the two following packages with the two lines below:

```r
install.packages("ggplot2")
install.packages("NLP")
```

II.   Usage

a. Running the app

To run the app, create a folder, and put the R file in it. It is important not to change the name of the R file. It must always be app.R. So that when you call the file, R knows exactly what to do with it. After this, you need to load the shiny library and use the runApp() command as showed just below:

```
library("shiny")
#Path is the path where the app.R file is.
runApp("Path", launch.browser = TRUE)
```

You need to be careful about one thing when you put in the path. You have to change "\" in "/". For instance:

```
runApp("C:\Users\Projet\Desktop\Internship\Application", launch.browser = TRUE)
```

will not work, but:

```
runApp("C:/Users/Projet/Internship_Visua_Combine", launch.browser = TRUE)
```

will work.

When you type this command line into the console, the application will open on your default browser. It could be open in R, but some functionalities don't work or badly work. However, if you want to open it in R, just remove launch.browser = TRUE from the command above.

b. Using the app

1. Overview

The first page of the app is called "Plot overview". It enables you to do several tasks:

- Choose a metric (for instance frequency) on the select input button.
- See an interactive plot, that changes according to the metric you choose
- See a data table, in which you can change the order of the data as you wish, find a word you are looking for…
- Select data points on the plot. In order to deselect them, you need to double-click on the plot. They will automatically appear on black in the data table, while the not selected points will appear on white in the data table.
- Upload a file. It is just here to see what the application will look like when you can actually upload a file on the app and process and visualize it. The back-end part is still in process, so this button is useless for now.
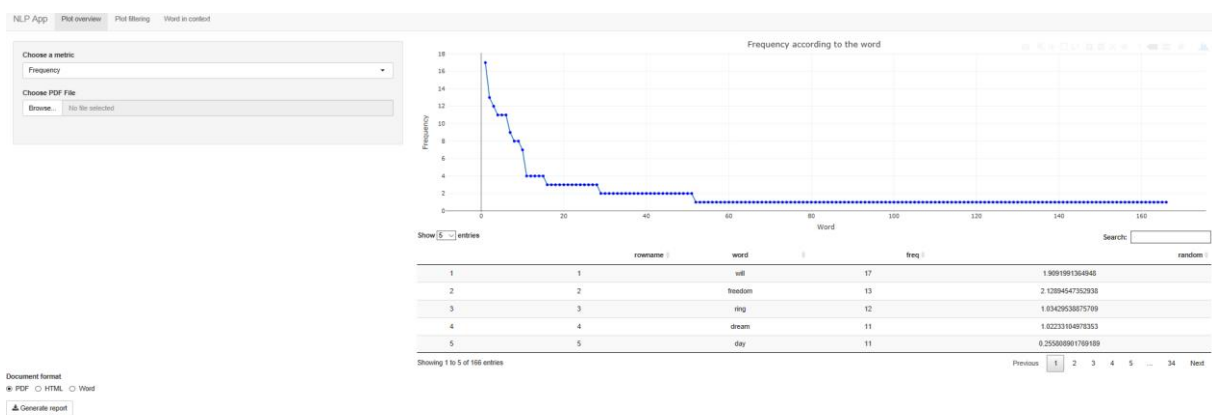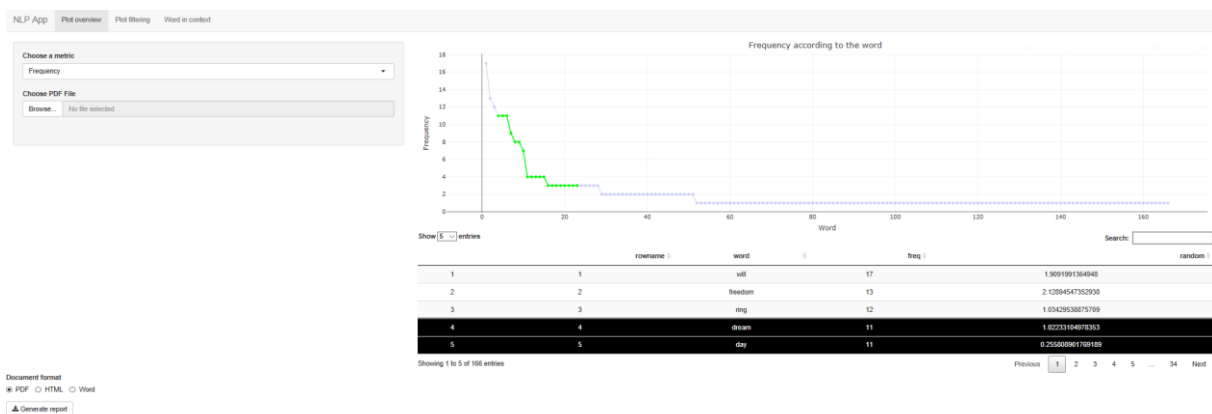

Figure 1: Screenshot of the overview

Figure 2: Screenshot of the overview after selection

2. Plot filtering

The second page of the app is called "Plot filtering". You can change the page with the panel that is at the top of the page.

This page enables you to:

- See a word cloud, with the words selected in the previous plot (the one in the "Plot overview" page).
- Choose a frequency range for the words in the cloud and a maximum number of words in the cloud with the slider inputs.
- Hover on the words and see the associated frequency
- Click on a word and see it appear at the bottom of the word cloud.
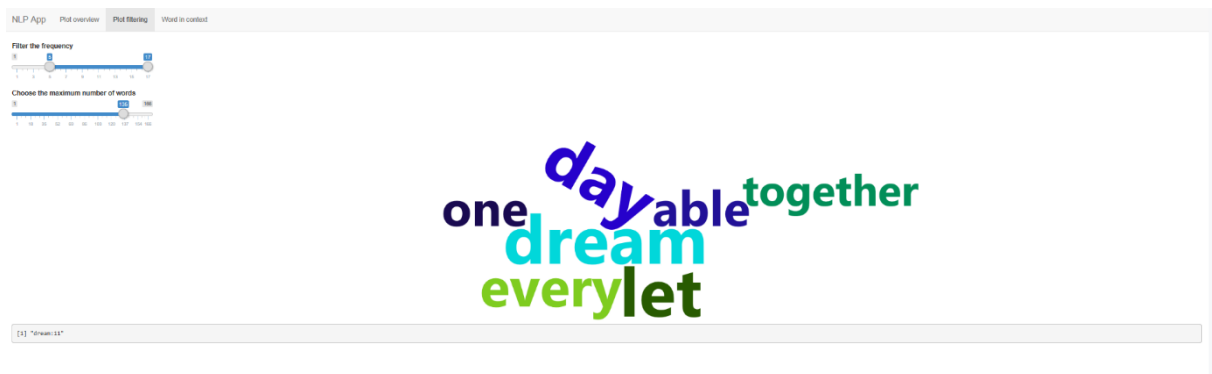

Figure 3: Screenshot of the filter

3. Report

Download your report. Indeed, the app enables the user to download a report of what he did during the session. This report includes the plots from the overview, the plots, table and wordcloud of the selected points. It is available in three output formats (PDF, HTML and WORD). It also contains a table of content.
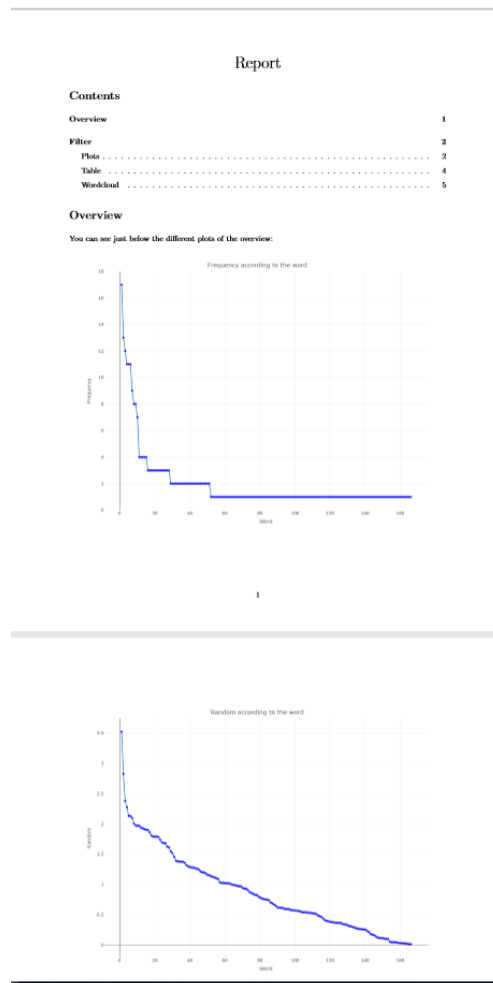
Figure 4: Screenshot of the beginning of a report

III.   Identified bugs and improvement

a. One word word cloud

The word cloud package used in the app doesn't support to have only word in it. So for instance, if you put the maximum number of words to one, nothing will appear, but if you put it to two, two words will indeed appear. This bug is being fixed.

b. Table options change

This bug appears in the "Plot overview" page in a particular situation. If you select the plot, then change the options of the data table, and deselect the plot, the options will not remain the same. So for instance if you select the plot, change the number of entries from 5 to 20 in the table, and deselect the plot, then the number of entries will be 5 instead of 20.

c. Plots

In the "Plot overview" page, there are two possible improvements for the plots:

- When selecting points on one plot, and then changing the metric, the points are no longer selected in the new plot with the new metric. This could be nice to have the points previously selected when changing the metric.
- Ordering the points according to the metric. Currently, the points are ordered by the the words' id, so they don't appear nice with metrics. There is an easy way to change that and adapt it to the data but it requires either a new column for each metric either a long computational process each time a plot is drawn, so there is a trade-off to do for this feature. The best option is probably the one that uses more space.

d. Wordcloud

There is another problem with the word cloud, but it is minor. Sometimes, when passing from the word cloud, and selecting a word in it, to the previous page ("Plot overview"), and then going back again at the word cloud without selecting words in the plot, the word cloud is empty (which is normal), but there is still the hover style from the word selected the last time.

Contact

If you have any question about the app, troubles to install/use it or would like to report bugs, please free to contact us.

Colette Voisembert   c.voisembert@gmail.com

Leon Migus   leonmigus@gmail.com