Spring Semester 2018
Iowa State University

LING 410X - Language as Data

Assignment 6
Submission Deadline: 14 April 2018 end of the day

Instructions: This assignment consists of only one question which carries
10 marks. Upload your submission as a Rmarkdown report (in html or doc
or pdf format) with the file name as your_Lastname_A6.html/pdf/docx.

Question 1

Take any two books (.txt files or any other format you want) from guten-
berg.org, follow whatever pre-processing you want. Using the wordcloud
package in R, do the following:

- Create a word cloud for each of these files.

- Create a commonality cloud for both the files together

- Create a comparison cloud for both files together.

Note: I want the report to be like a report. Not just presenting lines
and lines of R code.
References:

- Documentation of wordcloud package:
  `https://cran.r-project.org/web/packages/wordcloud/wordcloud.`
  `pdf`

- Here is a blogpost on this package, which gives you some useful infor-
  mation:
  `http://blog.fellstat.com/?cat=11`

- An old tutorial website on mining twitter with R that shows examples
  of commonality and comparison clouds :
  `https://sites.google.com/site/miningtwitter` - please note: This
  tutorial is old and it should only be used as an overview.