R for Sysadmins: WordClouds

Kwan Lowe

8/26/2016

R for SysAdmins: WordClouds

A wordcloud is a PHB-friendly way to visualize where the hot areas are in your problem ticket queue. More common words a printed in a larger font so it is easy to see systems that are generating tickets.

For this example, I exported my Remedy ticket queue to a CSV file. It required some basic massaging to clean up Microsoft encoding and drop the mail headers. That process, unfortunately, is out of scope for this demo. :)

```
library(tm)
```

```
## Loading required package: NLP
```

```
library(SnowballC)
library(wordcloud)
```

```
## Loading required package: RColorBrewer
```

This example is mostly verbatim from the R-bloggers example in the resources section. Main addition was to add some custom removed words.

```
mywords <- c("Kwan", "Lowe", "Management" )
remedy <- read.csv("remedy.csv", stringsAsFactors = FALSE)

remedyCorpus <- Corpus(VectorSource(remedy$Text))
remedyCorpus <- tm_map(remedyCorpus, PlainTextDocument)
remedyCorpus <- tm_map(remedyCorpus, removePunctuation)
remedyCorpus <- tm_map(remedyCorpus, removeWords, stopwords('english') )
remedyCorpus <- tm_map(remedyCorpus, removeWords, mywords )
wordcloud(remedyCorpus, max.words = 80, random.order = FALSE)
```

It may take a few minutes to run depending on the size of your dataset.

My example datafile is somewhat noisy. If you wanted to use this process in a production environment, you might consider using error codes or other terms.

References

[R-bloggers.com Article][https://www.r-bloggers.com/building-wordclouds-in-r/]

End.