# What Next?

- Feature Engineering:
  - How about tri-grams, 4-grams, etc.?
  - We engineered TextLength – are there other features as well?

- Algorithms - we leveraged a mighty random forest, but could other algorithms do more with the data?
  - Boosted decision trees?
  - Support vector machines (SVM)?

- Learn more ways to analyze, understand, and work with text!

datasciencedojo
data science for everyone

Start Here – The Basics
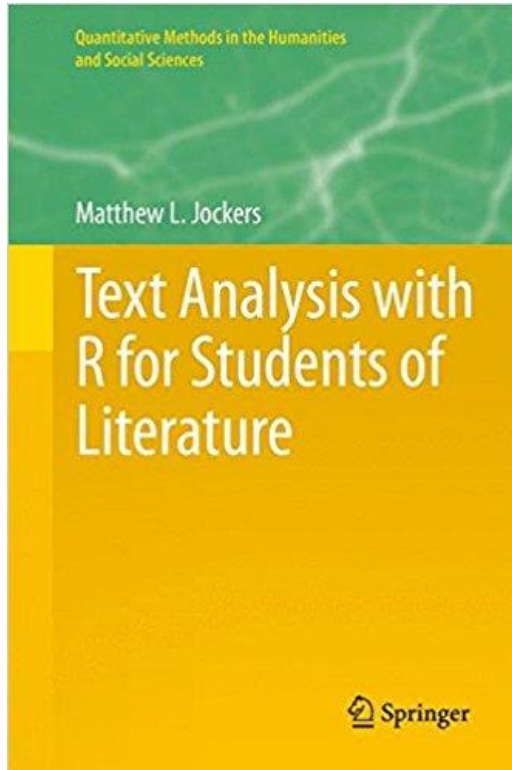
- Best introduction to thinking analytically about text.

- Accessible to a very broad audience.

- Illustrated many techniques not in series (e.g., topic modeling).

Next Stop – Java!?!?
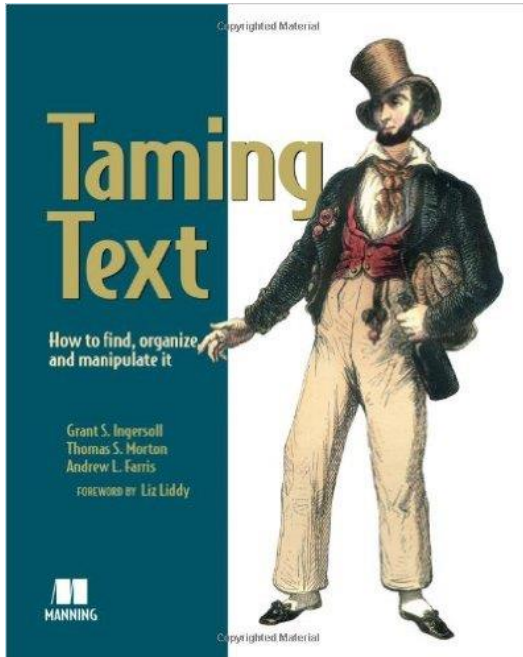
- Written for Java developers.

- However, a lot of theory easily implemented with R packages.

- Coverage of the Java OpenNLP library which has an R package!
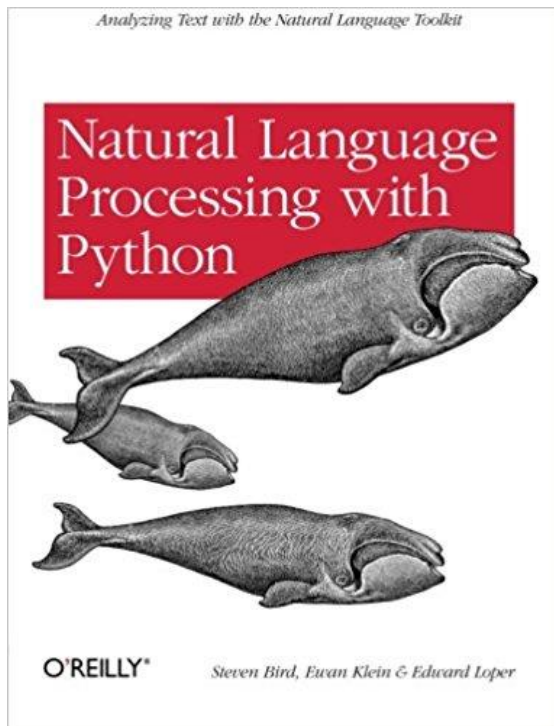
Survey the R Text Universe

- Many, many packages in R support text analytics.

- After studying "Taming Text" you'll know a lot more about what packages your need for what problem.

- Check out the CRAN NLP Task View for all the tasty R goodness!

Search == Text Analytics

- Much of information retrieval relates to text analytics.

- This is the single best resource on IR for beginners.

- Book is available for free online from Stanford!

Last Stop – Python!?!?

Analyzing Text with the Natural Language Toolkit

Natural Language Processing with Python

O'REILLY®

Steven Bird, Ewan Klein & Edward Loper

- Python has the Natural Language Toolkit (NLTK) – pure awesomeness.

- As with "Taming Text", this book contains a lot of theory to take your skills to the next level.

- Everything you can do with the NLTK you can do with R packages.

datasciencedojo

data science for everyone