AADITYA JAMUAR/ tan tana tan tan tan tara 1234

1. Loading a Corpus

Doc

Pdf

Txt

Others

1. Exploring the Corpus
2. Preparing the Corpus

Replace \/", used sometimes to separate alternative words, with a space.

Conversion to Lower Case

Remove Numbers

Remove Punctuation

Remove English Stop Words (for, very, and, of, are, etc, are common stop words)

Remove Own Stop Words

Strip Whitespace

Specific Transformations

Stemming (removing “es”, “ed”, “s”)

1. Creating a Document Term Matrix
2. Exploring the Document Term Matrix
3. Distribution of Term Frequencies
4. Conversion to Matrix and Save to CSV
5. Removing Sparse Terms
6. Identifying Frequent Items and Associations
7. Correlations Plots
8. Correlations Plot|Options
9. Plotting Word Frequencies
10. Word Clouds

Reducing Clutter With Max Words

Reducing Clutter With Min Freq

Adding Some Colour

Varying the Scaling

Rotating Words

1. Quantitative Analysis of Text

Word Length Counts

Letter Frequency

Letter and Position Heatmap

Miscellaneous Functions