

Heart of Darkness - Sentiment Analysis (Task 2)

Zara Waheed

Introduction

This report is a sentiment analysis using TrueNumbers of the novel Heart of Darkness by Joseph Conrad. Different lexicons are used to see the overall sentiment of the book.

Initial steps

Similar to Task 2, I did some cleaning of the text file of the book in order to make easier to work with. After downloading the book, I manually removed the table of contents. I also marked all the chapters so they were easier for TrueNumbers to read.

I then authorized the mssp server to get access. I set the test space to test3 and read lines from my book. I then used the Book2Tn file to ingest my book into TrueNumbers. After a couple of tries I got it to work. I then checked to see if it had been uploaded correctly and began to run queries on it.

To align with my initial analysis, three different lexicons were used in this analysis namely, “Afinn”, “Bing”, and “Nrc”.

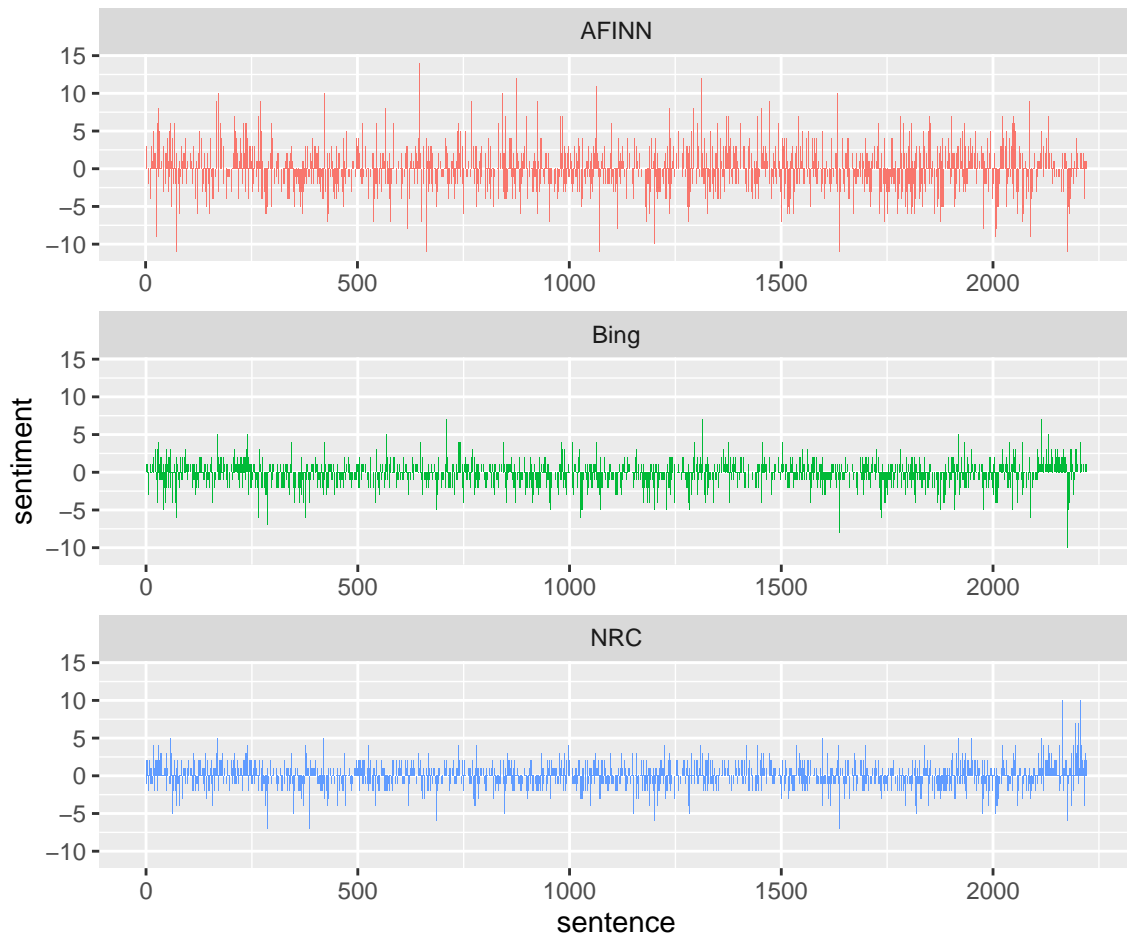
Data cleaning

The data cleaning was more of a challenge in this task. Creating data frames that contained sections for each chapter, paragraph and sentence required some extraction. In the end I ended up with a data frame that had the relevant divisions as shown below.

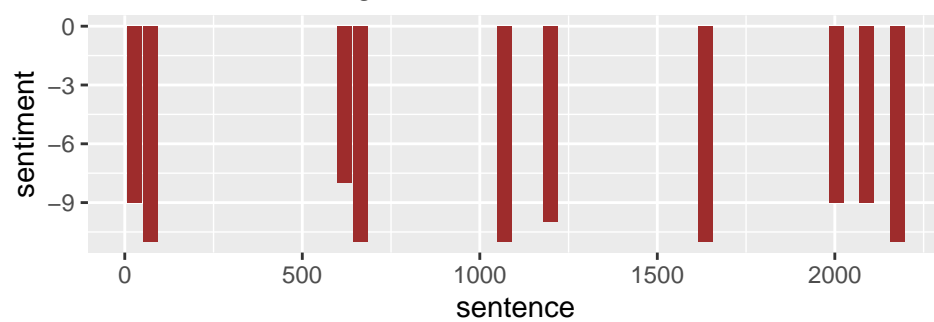
##	word	numpara	sentence
## 1	the	1	1
## 2	nellie	1	1
## 3	a	1	1
## 4	cruising	1	1
## 5	yawl	1	1
## 6	swung	1	1

I explored the same three lexicons as Task 2. The analysis is done at sentence and paragraph level.

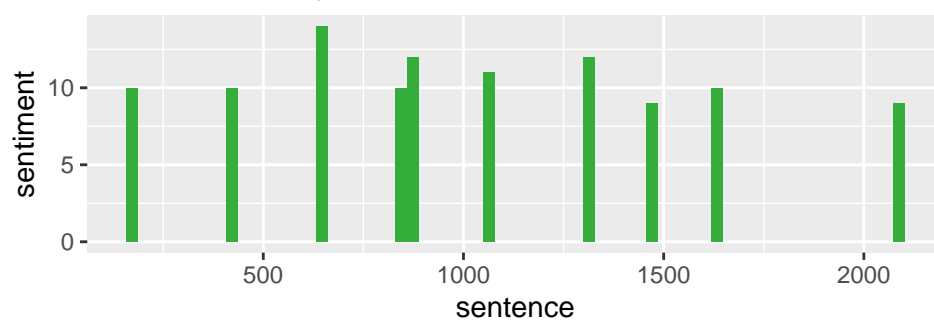
Sentence level Sentiment Analysis



AFINN – Most negative sentences



AFINN – Most positive sentences



Paragraph level Sentiment Analysis

