

FIT3152 Data analytics

Assignment 3

Name: Rhyme Bulbul

Student ID: 31865224

AI statement:

Task 1

I gathered a collection of 19 text documents from different sources that covered a range of themes for this task. Blogs, news stories, movie reviews, and more are included in these publications. The Appendix contains the list of references for these documents.

Text files were created by copying and pasting the document contents.

Task 2

Task 1 already had the document's contents pasted into text files. The text files were assembled into a single docs folder within the working directory in order to generate the corpus. Next, the code that follows is executed. It imports the necessary libraries, sets the seed (to ensure graph reproducibility), and uses the Corpus() function from the tm package to construct the corpus.

```
rm(list = ls())
set.seed(31865224)
#install.packages("tm")
library(tm)
#install.packages("cluster")
library(cluster)
#install.packages("igraph")
library(igraph)

cname <- file.path(".", "text")
docs <- Corpus(DirSource((cname)))
list.files("text")
```

```
## [1] "ai01.txt"          "australia01.txt"  "catchme01.txt"   "deadpool01.txt"
## [5] "fast01.txt"        "fast02.txt"      "impossible01.txt" "johnwick01.txt"
## [9] "linux01.txt"       "linux02.txt"     "linux03.txt"     "linux04.txt"
## [13] "pirate01.txt"      "pirate02.txt"    "prestige01.txt"  "shrek01.txt"
## [17] "starwars01.txt"    "terminator01.txt" "wallstreet01.txt"
```

Documents are named based on the topic, suffixed by a number. Such as, linux01.txt is the first text document on linux systems, followed by linux02.txt the second text document on the same topic, and so fourth.

Task 3

Appendix

References

Wikimedia Foundation. (2024, May 14). Ubuntu. Wikipedia. <https://en.wikipedia.org/wiki/Ubuntu>

Wikimedia Foundation. (2024c, May 18). Linux kernel. Wikipedia. https://en.wikipedia.org/wiki/Linux_kernel

Wikimedia Foundation. (2024a, April 5). LineageOS. Wikipedia. <https://en.wikipedia.org/wiki/LineageOS>

Wikimedia Foundation. (2024b, May 7). Android (Operating System). Wikipedia. [https://en.wikipedia.org/wiki/Android_\(operating_system\)](https://en.wikipedia.org/wiki/Android_(operating_system))

Encyclopædia Britannica, inc. (2024, May 14). Australia. Encyclopædia Britannica. <https://www.britannica.com/place/Australia>

Goode, L. (2024, May 14). It's the end of google search as we know it. Wired. <https://www.wired.com/story/google-io-end-of-google-search/>

IMDb.com. (2003, June 6). 2 fast 2 furious. IMDb. https://www.imdb.com/title/tt0322259/?ref_=nv_sr_s_rsg_0_tt_7_nm_1_q_2%2520fast

IMDb.com. (2009, April 3). Fast & Furious. IMDb. https://www.imdb.com/title/tt1013752/?ref_=nv_sr_s_rsg_1_tt_7_nm_0_q_fast%2520and%2520fur

IMDb.com. (2003b, July 9). Pirates of the Caribbean: The curse of the black pearl. IMDb. https://www.imdb.com/title/tt0325980/?ref_=nv_sr_s_rsg_0_tt_8_nm_0_q_pirates

IMDb.com. (2017, May 26). Pirates of the caribbean: Dead men tell no tales. IMDb. https://www.imdb.com/title/tt1790809/?ref_=nv_sr_s_rsg_3_tt_8_nm_0_q_pirates

IMDb.com. (2014, October 24). John Wick. IMDb. https://www.imdb.com/title/tt2911666/?ref_=nv_sr_s_rsg_1_tt_7_nm_1_q_jon%2520wick

IMDb.com. (1996, May 22). Mission: Impossible. IMDb. https://www.imdb.com/title/tt0117060/?ref_=nv_sr_s_rsg_0_tt_8_nm_0_q_mission

IMDb.com. (2016, February 12). Deadpool. IMDb. https://www.imdb.com/title/tt1431045/?ref_=nv_sr_s_rsg_3_tt_6_nm_2_q_deadpool

IMDb.com. (1984, October 26). The terminator. IMDb. https://www.imdb.com/title/tt0088247/?ref_=nv_sr_s_rsg_0_tt_8_nm_0_q_terminator

IMDb.com. (1977, May 25). Star wars: Episode IV - A new hope. IMDb. https://www.imdb.com/title/tt0076759/?ref_=nv_sr_s_rsg_4_tt_7_nm_0_q_star%2520wars

IMDb.com. (2004, May 19). Shrek 2. IMDb. https://www.imdb.com/title/tt0298148/?ref_=nv_sr_s_rsg_0_tt_8_nm_0_q_shrek%25202

IMDb.com. (2006, October 20). The prestige. IMDb. https://www.imdb.com/title/tt0482571/?ref_=nv_sr_s_rsg_0_tt_8_nm_0_q_prestige

IMDb.com. (2013, December 25). The wolf of wall street. IMDb. https://www.imdb.com/title/tt0993846/?ref_=nm_filmg_t_38_prd

IMDb.com. (2002, December 25). Catch me if you can. IMDb. https://www.imdb.com/title/tt0264464/?ref_=nv_sr_s_rsg_0_tt_8_nm_0_q_catch%2520me