

# Markdown\_Week8\_Week9

Aymeric Collart

2025-10-02

## 1. Prepare the environment

```
library(rvest)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(xml2)
library(openxlsx)
library(stringr)
```

## 2. Scraping

### 2.1 List the URLs of the Chinese classic books

```
PageUrl <- "https://millionbook.net/gd/gdxs.html"
PageClassChinLit <- read_html(PageUrl)

ListBooks <- PageClassChinLit %>%
  html_nodes(".tt1") %>%
  html_text()

ListBooks <- ListBooks[-c(90)] # Remove blank one

ListURL <- PageClassChinLit %>%
  html_nodes(".tt1") %>%
```

```

html_nodes("a") %>%
html_attr("href")

ListBooks <- as.data.frame(ListBooks)
ListBooks$ListURL <- paste0("https://millionbook.net/gd/", ListURL)

## Check if books finish with "index.html"
ListBooks$has_index <- str_detect(ListBooks$ListURL, "index.html")

## Remove books without index (other scraping process)
ListBooks$has_index <- as.factor(ListBooks$has_index)
ListBooks <- droplevels(subset(ListBooks, has_index == TRUE))
rownames(ListBooks) <- NULL

```

## 2.2. Retrieve the URLs of the chapters of a book

### 2.2.1 One page to test

```

IndBook <- read_html(ListBooks$ListURL[81])

ChapterURL <- IndBook %>%
  html_nodes("td") %>%
  html_nodes("a") %>%
  html_attr("href")

ChapterURL <- ChapterURL[-c(1, 2, 3)]
print(ChapterURL)

```

```

## [1] "000.htm" "001.htm" "002.htm" "003.htm" "004.htm" "005.htm" "006.htm"
## [8] "007.htm" "008.htm" "009.htm" "010.htm" "011.htm" "012.htm" "013.htm"
## [15] "014.htm" "015.htm" "016.htm" "017.htm" "018.htm" "019.htm" "020.htm"
## [22] "021.htm" "022.htm" "023.htm" "024.htm" "025.htm" "026.htm" "027.htm"
## [29] "028.htm" "029.htm" "030.htm" "031.htm" "032.htm" "033.htm" "034.htm"
## [36] "035.htm" "036.htm" "037.htm" "038.htm" "039.htm" "040.htm" "041.htm"
## [43] "042.htm" "043.htm" "044.htm" "045.htm" "046.htm" "047.htm" "048.htm"
## [50] "049.htm" "050.htm" "051.htm" "052.htm" "053.htm" "054.htm" "055.htm"
## [57] "056.htm" "057.htm" "058.htm" "059.htm" "060.htm" "061.htm" "062.htm"
## [64] "063.htm" "064.htm" "065.htm" "066.htm" "067.htm" "068.htm" "069.htm"
## [71] "070.htm" "071.htm" "072.htm" "073.htm" "074.htm" "075.htm" "076.htm"
## [78] "077.htm" "078.htm" "079.htm" "080.htm"

```

```

ChapterURL <- unique(ChapterURL)
print(ChapterURL)

```

```

## [1] "000.htm" "001.htm" "002.htm" "003.htm" "004.htm" "005.htm" "006.htm"
## [8] "007.htm" "008.htm" "009.htm" "010.htm" "011.htm" "012.htm" "013.htm"
## [15] "014.htm" "015.htm" "016.htm" "017.htm" "018.htm" "019.htm" "020.htm"
## [22] "021.htm" "022.htm" "023.htm" "024.htm" "025.htm" "026.htm" "027.htm"
## [29] "028.htm" "029.htm" "030.htm" "031.htm" "032.htm" "033.htm" "034.htm"
## [36] "035.htm" "036.htm" "037.htm" "038.htm" "039.htm" "040.htm" "041.htm"

```

```
## [43] "042.htm" "043.htm" "044.htm" "045.htm" "046.htm" "047.htm" "048.htm"
## [50] "049.htm" "050.htm" "051.htm" "052.htm" "053.htm" "054.htm" "055.htm"
## [57] "056.htm" "057.htm" "058.htm" "059.htm" "060.htm" "061.htm" "062.htm"
## [64] "063.htm" "064.htm" "065.htm" "066.htm" "067.htm" "068.htm" "069.htm"
## [71] "070.htm" "071.htm" "072.htm" "073.htm" "074.htm" "075.htm" "076.htm"
## [78] "077.htm" "078.htm" "079.htm" "080.htm"
```

```
## Reconstruct the full URL
```

```
TempBookURL <- gsub("index.html", "", ListBooks$ListURL[81])
FullChapterURL <- paste0(TempBookURL, ChapterURL)
print(FullChapterURL)
```

```
## [1] "https://millionbook.net/gd/q/qiancai/syqz/000.htm"
## [2] "https://millionbook.net/gd/q/qiancai/syqz/001.htm"
## [3] "https://millionbook.net/gd/q/qiancai/syqz/002.htm"
## [4] "https://millionbook.net/gd/q/qiancai/syqz/003.htm"
## [5] "https://millionbook.net/gd/q/qiancai/syqz/004.htm"
## [6] "https://millionbook.net/gd/q/qiancai/syqz/005.htm"
## [7] "https://millionbook.net/gd/q/qiancai/syqz/006.htm"
## [8] "https://millionbook.net/gd/q/qiancai/syqz/007.htm"
## [9] "https://millionbook.net/gd/q/qiancai/syqz/008.htm"
## [10] "https://millionbook.net/gd/q/qiancai/syqz/009.htm"
## [11] "https://millionbook.net/gd/q/qiancai/syqz/010.htm"
## [12] "https://millionbook.net/gd/q/qiancai/syqz/011.htm"
## [13] "https://millionbook.net/gd/q/qiancai/syqz/012.htm"
## [14] "https://millionbook.net/gd/q/qiancai/syqz/013.htm"
## [15] "https://millionbook.net/gd/q/qiancai/syqz/014.htm"
## [16] "https://millionbook.net/gd/q/qiancai/syqz/015.htm"
## [17] "https://millionbook.net/gd/q/qiancai/syqz/016.htm"
## [18] "https://millionbook.net/gd/q/qiancai/syqz/017.htm"
## [19] "https://millionbook.net/gd/q/qiancai/syqz/018.htm"
## [20] "https://millionbook.net/gd/q/qiancai/syqz/019.htm"
## [21] "https://millionbook.net/gd/q/qiancai/syqz/020.htm"
## [22] "https://millionbook.net/gd/q/qiancai/syqz/021.htm"
## [23] "https://millionbook.net/gd/q/qiancai/syqz/022.htm"
## [24] "https://millionbook.net/gd/q/qiancai/syqz/023.htm"
## [25] "https://millionbook.net/gd/q/qiancai/syqz/024.htm"
## [26] "https://millionbook.net/gd/q/qiancai/syqz/025.htm"
## [27] "https://millionbook.net/gd/q/qiancai/syqz/026.htm"
## [28] "https://millionbook.net/gd/q/qiancai/syqz/027.htm"
## [29] "https://millionbook.net/gd/q/qiancai/syqz/028.htm"
## [30] "https://millionbook.net/gd/q/qiancai/syqz/029.htm"
## [31] "https://millionbook.net/gd/q/qiancai/syqz/030.htm"
## [32] "https://millionbook.net/gd/q/qiancai/syqz/031.htm"
## [33] "https://millionbook.net/gd/q/qiancai/syqz/032.htm"
## [34] "https://millionbook.net/gd/q/qiancai/syqz/033.htm"
## [35] "https://millionbook.net/gd/q/qiancai/syqz/034.htm"
## [36] "https://millionbook.net/gd/q/qiancai/syqz/035.htm"
## [37] "https://millionbook.net/gd/q/qiancai/syqz/036.htm"
## [38] "https://millionbook.net/gd/q/qiancai/syqz/037.htm"
## [39] "https://millionbook.net/gd/q/qiancai/syqz/038.htm"
## [40] "https://millionbook.net/gd/q/qiancai/syqz/039.htm"
## [41] "https://millionbook.net/gd/q/qiancai/syqz/040.htm"
```

```
## [42] "https://millionbook.net/gd/q/qiancai/syqz/041.htm"
## [43] "https://millionbook.net/gd/q/qiancai/syqz/042.htm"
## [44] "https://millionbook.net/gd/q/qiancai/syqz/043.htm"
## [45] "https://millionbook.net/gd/q/qiancai/syqz/044.htm"
## [46] "https://millionbook.net/gd/q/qiancai/syqz/045.htm"
## [47] "https://millionbook.net/gd/q/qiancai/syqz/046.htm"
## [48] "https://millionbook.net/gd/q/qiancai/syqz/047.htm"
## [49] "https://millionbook.net/gd/q/qiancai/syqz/048.htm"
## [50] "https://millionbook.net/gd/q/qiancai/syqz/049.htm"
## [51] "https://millionbook.net/gd/q/qiancai/syqz/050.htm"
## [52] "https://millionbook.net/gd/q/qiancai/syqz/051.htm"
## [53] "https://millionbook.net/gd/q/qiancai/syqz/052.htm"
## [54] "https://millionbook.net/gd/q/qiancai/syqz/053.htm"
## [55] "https://millionbook.net/gd/q/qiancai/syqz/054.htm"
## [56] "https://millionbook.net/gd/q/qiancai/syqz/055.htm"
## [57] "https://millionbook.net/gd/q/qiancai/syqz/056.htm"
## [58] "https://millionbook.net/gd/q/qiancai/syqz/057.htm"
## [59] "https://millionbook.net/gd/q/qiancai/syqz/058.htm"
## [60] "https://millionbook.net/gd/q/qiancai/syqz/059.htm"
## [61] "https://millionbook.net/gd/q/qiancai/syqz/060.htm"
## [62] "https://millionbook.net/gd/q/qiancai/syqz/061.htm"
## [63] "https://millionbook.net/gd/q/qiancai/syqz/062.htm"
## [64] "https://millionbook.net/gd/q/qiancai/syqz/063.htm"
## [65] "https://millionbook.net/gd/q/qiancai/syqz/064.htm"
## [66] "https://millionbook.net/gd/q/qiancai/syqz/065.htm"
## [67] "https://millionbook.net/gd/q/qiancai/syqz/066.htm"
## [68] "https://millionbook.net/gd/q/qiancai/syqz/067.htm"
## [69] "https://millionbook.net/gd/q/qiancai/syqz/068.htm"
## [70] "https://millionbook.net/gd/q/qiancai/syqz/069.htm"
## [71] "https://millionbook.net/gd/q/qiancai/syqz/070.htm"
## [72] "https://millionbook.net/gd/q/qiancai/syqz/071.htm"
## [73] "https://millionbook.net/gd/q/qiancai/syqz/072.htm"
## [74] "https://millionbook.net/gd/q/qiancai/syqz/073.htm"
## [75] "https://millionbook.net/gd/q/qiancai/syqz/074.htm"
## [76] "https://millionbook.net/gd/q/qiancai/syqz/075.htm"
## [77] "https://millionbook.net/gd/q/qiancai/syqz/076.htm"
## [78] "https://millionbook.net/gd/q/qiancai/syqz/077.htm"
## [79] "https://millionbook.net/gd/q/qiancai/syqz/078.htm"
## [80] "https://millionbook.net/gd/q/qiancai/syqz/079.htm"
## [81] "https://millionbook.net/gd/q/qiancai/syqz/080.htm"
```

## 2.2.2 Find the URLs of the chapters from all the books

```
## 3 to 4 minutes
URL_BookChapters_total <- data.frame(matrix(ncol = 1, nrow = 0))

for (i in 1:length(ListBooks$ListURL)){
  IndBook <- read_html(ListBooks$ListURL[i])

  ChapterURL <- IndBook %>%
    html_nodes("td") %>%
    html_nodes("a") %>%
```

```

    html_attr("href")

ChapterURL <- ChapterURL[-c(1, 2, 3)]
ChapterURL <- unique(ChapterURL)

TempBookURL <- gsub("index.html", "", ListBooks$ListURL[i])
FullChapterURL <- paste0(TempBookURL, ChapterURL)

URL_BookChapters <- (data.frame(ChapterURL = FullChapterURL,
                                Book = ListBooks$ListBooks[i]))

URL_BookChapters_total <- rbind(URL_BookChapters_total, URL_BookChapters)
}

## Remove links to images (.jpg)
## Check if books finish with "index.html"
URL_BookChapters_total$image <- str_detect(URL_BookChapters_total$ChapterURL, ".jpg")
URL_BookChapters_total$image <- as.factor(URL_BookChapters_total$image)
URL_BookChapters_total <- droplevels(subset(URL_BookChapters_total, image != TRUE))
rownames(URL_BookChapters_total) <- NULL

## Problems occurred:
## 1- .jpg documents
## 2- Inconsistent structure of the webpage
## - 聊齋志異 (蒲松齡)
## - 羅貫中文集
## - 岳全傳 (錢彩)

```

## 2.3 Scrape the content of the chapters

### 2.3.1 Test with one page

```

Chapter <- data.frame()

j = 3672

OneChapter <- read_html(URL_BookChapters_total$ChapterURL[j])

ChapterContent <- OneChapter %>%
  html_nodes(".tt2") %>%
  html_text()

ChapterTitle <- sub("\\r.*", "", ChapterContent)

Chapter <- (data.frame(Book = URL_BookChapters_total$Book[j],
                      Title = ChapterTitle,
                      Content = ChapterContent,
                      URL = URL_BookChapters_total$ChapterURL[j]))

head(Chapter)

```

##	Book	Title
----	------	-------

```
## 1 鬼谷四友志(楊景脩) 卷之一下 忌刻小人行毒計 忠直良友詐風魔
##
## 1 卷之一下 忌刻小人行毒計 忠直良友詐風魔\r\n\r\n 蘇秦、張儀于鬼谷先生處學游，這日看見魏王
```

### 2.3.2 Scrape all the chapters for one book

```
Book_total <- data.frame()

for (j in 1:length(URL_BookChapters_total$ChapterURL)){
  tryCatch({

    OneChapter <- read_html(URL_BookChapters_total$ChapterURL[j])

    ChapterContent <- OneChapter %>%
      html_nodes(".tt2") %>%
      html_text()

    ChapterTitle <- sub("\r.*", "", ChapterContent)

    Chapter <- (data.frame(Book = URL_BookChapters_total$Book[j],
                          Title = ChapterTitle,
                          Content = ChapterContent,
                          URL = URL_BookChapters_total$ChapterURL[j]))
    Book_total <- rbind(Book_total, Chapter)

    # Update the progress
    percentage_completed <- round(j/length(URL_BookChapters_total$ChapterURL) * 100, 2)
    cat(c("Process", j, "out of", length(URL_BookChapters_total$ChapterURL), "percentage
    ↵ completed:", percentage_completed, "%", "\n"))

  }, error=function(e){cat("ERROR :",conditionMessage(e), "\n")})
}
```

```
## Process 1 out of 4556 percentage completed: 0.02 %
## Process 2 out of 4556 percentage completed: 0.04 %
## Process 3 out of 4556 percentage completed: 0.07 %
## Process 4 out of 4556 percentage completed: 0.09 %
## Process 5 out of 4556 percentage completed: 0.11 %
## Process 6 out of 4556 percentage completed: 0.13 %
## Process 7 out of 4556 percentage completed: 0.15 %
## Process 8 out of 4556 percentage completed: 0.18 %
## Process 9 out of 4556 percentage completed: 0.2 %
## Process 10 out of 4556 percentage completed: 0.22 %
## Process 11 out of 4556 percentage completed: 0.24 %
## Process 12 out of 4556 percentage completed: 0.26 %
## Process 13 out of 4556 percentage completed: 0.29 %
## Process 14 out of 4556 percentage completed: 0.31 %
## Process 15 out of 4556 percentage completed: 0.33 %
## Process 16 out of 4556 percentage completed: 0.35 %
```

## Process 17 out of 4556 percentage completed: 0.37 %  
## Process 18 out of 4556 percentage completed: 0.4 %  
## Process 19 out of 4556 percentage completed: 0.42 %  
## Process 20 out of 4556 percentage completed: 0.44 %  
## Process 21 out of 4556 percentage completed: 0.46 %  
## Process 22 out of 4556 percentage completed: 0.48 %  
## Process 23 out of 4556 percentage completed: 0.5 %  
## Process 24 out of 4556 percentage completed: 0.53 %  
## Process 25 out of 4556 percentage completed: 0.55 %  
## Process 26 out of 4556 percentage completed: 0.57 %  
## Process 27 out of 4556 percentage completed: 0.59 %  
## Process 28 out of 4556 percentage completed: 0.61 %  
## Process 29 out of 4556 percentage completed: 0.64 %  
## Process 30 out of 4556 percentage completed: 0.66 %  
## Process 31 out of 4556 percentage completed: 0.68 %  
## Process 32 out of 4556 percentage completed: 0.7 %  
## Process 33 out of 4556 percentage completed: 0.72 %  
## Process 34 out of 4556 percentage completed: 0.75 %  
## Process 35 out of 4556 percentage completed: 0.77 %  
## Process 36 out of 4556 percentage completed: 0.79 %  
## Process 37 out of 4556 percentage completed: 0.81 %  
## Process 38 out of 4556 percentage completed: 0.83 %  
## Process 39 out of 4556 percentage completed: 0.86 %  
## Process 40 out of 4556 percentage completed: 0.88 %  
## Process 41 out of 4556 percentage completed: 0.9 %  
## Process 42 out of 4556 percentage completed: 0.92 %  
## Process 43 out of 4556 percentage completed: 0.94 %  
## Process 44 out of 4556 percentage completed: 0.97 %  
## Process 45 out of 4556 percentage completed: 0.99 %  
## Process 46 out of 4556 percentage completed: 1.01 %  
## Process 47 out of 4556 percentage completed: 1.03 %  
## Process 48 out of 4556 percentage completed: 1.05 %  
## Process 49 out of 4556 percentage completed: 1.08 %  
## Process 50 out of 4556 percentage completed: 1.1 %  
## Process 51 out of 4556 percentage completed: 1.12 %  
## Process 52 out of 4556 percentage completed: 1.14 %  
## Process 53 out of 4556 percentage completed: 1.16 %  
## Process 54 out of 4556 percentage completed: 1.19 %  
## Process 55 out of 4556 percentage completed: 1.21 %  
## Process 56 out of 4556 percentage completed: 1.23 %  
## Process 57 out of 4556 percentage completed: 1.25 %  
## Process 58 out of 4556 percentage completed: 1.27 %  
## Process 59 out of 4556 percentage completed: 1.29 %  
## Process 60 out of 4556 percentage completed: 1.32 %  
## Process 61 out of 4556 percentage completed: 1.34 %  
## Process 62 out of 4556 percentage completed: 1.36 %  
## Process 63 out of 4556 percentage completed: 1.38 %  
## Process 64 out of 4556 percentage completed: 1.4 %  
## Process 65 out of 4556 percentage completed: 1.43 %  
## Process 66 out of 4556 percentage completed: 1.45 %  
## Process 67 out of 4556 percentage completed: 1.47 %  
## Process 68 out of 4556 percentage completed: 1.49 %  
## Process 69 out of 4556 percentage completed: 1.51 %  
## Process 70 out of 4556 percentage completed: 1.54 %

```

## Process 71 out of 4556 percentage completed: 1.56 %
## Process 72 out of 4556 percentage completed: 1.58 %
## Process 73 out of 4556 percentage completed: 1.6 %
## Process 74 out of 4556 percentage completed: 1.62 %
## Process 75 out of 4556 percentage completed: 1.65 %
## Process 76 out of 4556 percentage completed: 1.67 %
## Process 77 out of 4556 percentage completed: 1.69 %
## Process 78 out of 4556 percentage completed: 1.71 %
## Process 79 out of 4556 percentage completed: 1.73 %
## Process 80 out of 4556 percentage completed: 1.76 %
## Process 81 out of 4556 percentage completed: 1.78 %
## Process 82 out of 4556 percentage completed: 1.8 %
## Process 83 out of 4556 percentage completed: 1.82 %
## Process 84 out of 4556 percentage completed: 1.84 %
## Process 85 out of 4556 percentage completed: 1.87 %
## Process 86 out of 4556 percentage completed: 1.89 %
## Process 87 out of 4556 percentage completed: 1.91 %
## Process 88 out of 4556 percentage completed: 1.93 %
## Process 89 out of 4556 percentage completed: 1.95 %
## Process 90 out of 4556 percentage completed: 1.98 %
## Process 91 out of 4556 percentage completed: 2 %
## Process 92 out of 4556 percentage completed: 2.02 %
## Process 93 out of 4556 percentage completed: 2.04 %
## Process 94 out of 4556 percentage completed: 2.06 %
## Process 95 out of 4556 percentage completed: 2.09 %
## Process 96 out of 4556 percentage completed: 2.11 %
## Process 97 out of 4556 percentage completed: 2.13 %
## Process 98 out of 4556 percentage completed: 2.15 %
## Process 99 out of 4556 percentage completed: 2.17 %
## Process 100 out of 4556 percentage completed: 2.19 %
## Process 101 out of 4556 percentage completed: 2.22 %
## Process 102 out of 4556 percentage completed: 2.24 %
## Process 103 out of 4556 percentage completed: 2.26 %
## Process 104 out of 4556 percentage completed: 2.28 %
## Process 105 out of 4556 percentage completed: 2.3 %
## Process 106 out of 4556 percentage completed: 2.33 %
## Process 107 out of 4556 percentage completed: 2.35 %
## Process 108 out of 4556 percentage completed: 2.37 %
## Process 109 out of 4556 percentage completed: 2.39 %
## Process 110 out of 4556 percentage completed: 2.41 %
## Process 111 out of 4556 percentage completed: 2.44 %
## Process 112 out of 4556 percentage completed: 2.46 %
## Process 113 out of 4556 percentage completed: 2.48 %
## Process 114 out of 4556 percentage completed: 2.5 %
## Process 115 out of 4556 percentage completed: 2.52 %
## Process 116 out of 4556 percentage completed: 2.55 %
## Process 117 out of 4556 percentage completed: 2.57 %
## Process 118 out of 4556 percentage completed: 2.59 %
## Process 119 out of 4556 percentage completed: 2.61 %
## Process 120 out of 4556 percentage completed: 2.63 %
## ERROR : arguments imply differing number of rows: 1, 0
## ERROR : arguments imply differing number of rows: 1, 0
## ERROR : arguments imply differing number of rows: 1, 0
## Process 124 out of 4556 percentage completed: 2.72 %

```



## Process 125 out of 4556 percentage completed: 2.74 %  
## Process 126 out of 4556 percentage completed: 2.77 %  
## Process 127 out of 4556 percentage completed: 2.79 %  
## Process 128 out of 4556 percentage completed: 2.81 %  
## Process 129 out of 4556 percentage completed: 2.83 %  
## Process 130 out of 4556 percentage completed: 2.85 %  
## Process 131 out of 4556 percentage completed: 2.88 %  
## Process 132 out of 4556 percentage completed: 2.9 %  
## Process 133 out of 4556 percentage completed: 2.92 %  
## Process 134 out of 4556 percentage completed: 2.94 %  
## Process 135 out of 4556 percentage completed: 2.96 %  
## Process 136 out of 4556 percentage completed: 2.99 %  
## Process 137 out of 4556 percentage completed: 3.01 %  
## Process 138 out of 4556 percentage completed: 3.03 %  
## ERROR : cannot open the connection  
## Process 140 out of 4556 percentage completed: 3.07 %  
## Process 141 out of 4556 percentage completed: 3.09 %  
## Process 142 out of 4556 percentage completed: 3.12 %  
## Process 143 out of 4556 percentage completed: 3.14 %  
## Process 144 out of 4556 percentage completed: 3.16 %  
## Process 145 out of 4556 percentage completed: 3.18 %  
## Process 146 out of 4556 percentage completed: 3.2 %  
## Process 147 out of 4556 percentage completed: 3.23 %  
## Process 148 out of 4556 percentage completed: 3.25 %  
## Process 149 out of 4556 percentage completed: 3.27 %  
## Process 150 out of 4556 percentage completed: 3.29 %  
## Process 151 out of 4556 percentage completed: 3.31 %  
## Process 152 out of 4556 percentage completed: 3.34 %  
## Process 153 out of 4556 percentage completed: 3.36 %  
## Process 154 out of 4556 percentage completed: 3.38 %  
## Process 155 out of 4556 percentage completed: 3.4 %  
## Process 156 out of 4556 percentage completed: 3.42 %  
## Process 157 out of 4556 percentage completed: 3.45 %  
## Process 158 out of 4556 percentage completed: 3.47 %  
## Process 159 out of 4556 percentage completed: 3.49 %  
## Process 160 out of 4556 percentage completed: 3.51 %  
## Process 161 out of 4556 percentage completed: 3.53 %  
## Process 162 out of 4556 percentage completed: 3.56 %  
## Process 163 out of 4556 percentage completed: 3.58 %  
## Process 164 out of 4556 percentage completed: 3.6 %  
## Process 165 out of 4556 percentage completed: 3.62 %  
## Process 166 out of 4556 percentage completed: 3.64 %  
## Process 167 out of 4556 percentage completed: 3.67 %  
## Process 168 out of 4556 percentage completed: 3.69 %  
## Process 169 out of 4556 percentage completed: 3.71 %  
## Process 170 out of 4556 percentage completed: 3.73 %  
## Process 171 out of 4556 percentage completed: 3.75 %  
## Process 172 out of 4556 percentage completed: 3.78 %  
## Process 173 out of 4556 percentage completed: 3.8 %  
## Process 174 out of 4556 percentage completed: 3.82 %  
## Process 175 out of 4556 percentage completed: 3.84 %  
## Process 176 out of 4556 percentage completed: 3.86 %  
## Process 177 out of 4556 percentage completed: 3.88 %  
## Process 178 out of 4556 percentage completed: 3.91 %

```

## Process 179 out of 4556 percentage completed: 3.93 %
## Process 180 out of 4556 percentage completed: 3.95 %
## Process 181 out of 4556 percentage completed: 3.97 %
## Process 182 out of 4556 percentage completed: 3.99 %
## Process 183 out of 4556 percentage completed: 4.02 %
## Process 184 out of 4556 percentage completed: 4.04 %
## Process 185 out of 4556 percentage completed: 4.06 %
## Process 186 out of 4556 percentage completed: 4.08 %
## Process 187 out of 4556 percentage completed: 4.1 %
## Process 188 out of 4556 percentage completed: 4.13 %
## Process 189 out of 4556 percentage completed: 4.15 %
## Process 190 out of 4556 percentage completed: 4.17 %
## Process 191 out of 4556 percentage completed: 4.19 %
## Process 192 out of 4556 percentage completed: 4.21 %
## Process 193 out of 4556 percentage completed: 4.24 %
## Process 194 out of 4556 percentage completed: 4.26 %
## Process 195 out of 4556 percentage completed: 4.28 %
## Process 196 out of 4556 percentage completed: 4.3 %
## Process 197 out of 4556 percentage completed: 4.32 %
## Process 198 out of 4556 percentage completed: 4.35 %
## Process 199 out of 4556 percentage completed: 4.37 %
## Process 200 out of 4556 percentage completed: 4.39 %
## Process 201 out of 4556 percentage completed: 4.41 %
## Process 202 out of 4556 percentage completed: 4.43 %
## Process 203 out of 4556 percentage completed: 4.46 %
## Process 204 out of 4556 percentage completed: 4.48 %
## Process 205 out of 4556 percentage completed: 4.5 %
## Process 206 out of 4556 percentage completed: 4.52 %
## Process 207 out of 4556 percentage completed: 4.54 %
## Process 208 out of 4556 percentage completed: 4.57 %
## Process 209 out of 4556 percentage completed: 4.59 %
## Process 210 out of 4556 percentage completed: 4.61 %
## Process 211 out of 4556 percentage completed: 4.63 %
## ERROR : cannot open the connection
## Process 213 out of 4556 percentage completed: 4.68 %
## Process 214 out of 4556 percentage completed: 4.7 %
## Process 215 out of 4556 percentage completed: 4.72 %
## Process 216 out of 4556 percentage completed: 4.74 %
## Process 217 out of 4556 percentage completed: 4.76 %
## Process 218 out of 4556 percentage completed: 4.78 %
## Process 219 out of 4556 percentage completed: 4.81 %
## Process 220 out of 4556 percentage completed: 4.83 %
## Process 221 out of 4556 percentage completed: 4.85 %
## Process 222 out of 4556 percentage completed: 4.87 %
## Process 223 out of 4556 percentage completed: 4.89 %
## Process 224 out of 4556 percentage completed: 4.92 %
## Process 225 out of 4556 percentage completed: 4.94 %
## Process 226 out of 4556 percentage completed: 4.96 %
## Process 227 out of 4556 percentage completed: 4.98 %
## ERROR : cannot open the connection
## Process 229 out of 4556 percentage completed: 5.03 %
## Process 230 out of 4556 percentage completed: 5.05 %
## Process 231 out of 4556 percentage completed: 5.07 %
## Process 232 out of 4556 percentage completed: 5.09 %

```

```

## Process 233 out of 4556 percentage completed: 5.11 %
## Process 234 out of 4556 percentage completed: 5.14 %
## Process 235 out of 4556 percentage completed: 5.16 %
## Process 236 out of 4556 percentage completed: 5.18 %
## Process 237 out of 4556 percentage completed: 5.2 %
## Process 238 out of 4556 percentage completed: 5.22 %
## Process 239 out of 4556 percentage completed: 5.25 %
## Process 240 out of 4556 percentage completed: 5.27 %
## Process 241 out of 4556 percentage completed: 5.29 %
## Process 242 out of 4556 percentage completed: 5.31 %
## Process 243 out of 4556 percentage completed: 5.33 %
## Process 244 out of 4556 percentage completed: 5.36 %
## Process 245 out of 4556 percentage completed: 5.38 %
## Process 246 out of 4556 percentage completed: 5.4 %
## Process 247 out of 4556 percentage completed: 5.42 %
## Process 248 out of 4556 percentage completed: 5.44 %
## Process 249 out of 4556 percentage completed: 5.47 %
## Process 250 out of 4556 percentage completed: 5.49 %
## Process 251 out of 4556 percentage completed: 5.51 %
## Process 252 out of 4556 percentage completed: 5.53 %
## Process 253 out of 4556 percentage completed: 5.55 %
## Process 254 out of 4556 percentage completed: 5.58 %
## Process 255 out of 4556 percentage completed: 5.6 %
## Process 256 out of 4556 percentage completed: 5.62 %
## Process 257 out of 4556 percentage completed: 5.64 %
## Process 258 out of 4556 percentage completed: 5.66 %
## Process 259 out of 4556 percentage completed: 5.68 %
## Process 260 out of 4556 percentage completed: 5.71 %
## Process 261 out of 4556 percentage completed: 5.73 %
## Process 262 out of 4556 percentage completed: 5.75 %
## Process 263 out of 4556 percentage completed: 5.77 %
## ERROR : cannot open the connection
## Process 265 out of 4556 percentage completed: 5.82 %
## Process 266 out of 4556 percentage completed: 5.84 %
## Process 267 out of 4556 percentage completed: 5.86 %
## Process 268 out of 4556 percentage completed: 5.88 %
## Process 269 out of 4556 percentage completed: 5.9 %
## Process 270 out of 4556 percentage completed: 5.93 %
## Process 271 out of 4556 percentage completed: 5.95 %
## Process 272 out of 4556 percentage completed: 5.97 %
## Process 273 out of 4556 percentage completed: 5.99 %
## Process 274 out of 4556 percentage completed: 6.01 %
## Process 275 out of 4556 percentage completed: 6.04 %
## Process 276 out of 4556 percentage completed: 6.06 %
## ERROR : cannot open the connection
## Process 278 out of 4556 percentage completed: 6.1 %
## Process 279 out of 4556 percentage completed: 6.12 %
## Process 280 out of 4556 percentage completed: 6.15 %
## Process 281 out of 4556 percentage completed: 6.17 %
## Process 282 out of 4556 percentage completed: 6.19 %
## Process 283 out of 4556 percentage completed: 6.21 %
## Process 284 out of 4556 percentage completed: 6.23 %
## Process 285 out of 4556 percentage completed: 6.26 %
## Process 286 out of 4556 percentage completed: 6.28 %

```

[illegible]

[illegible]

## Process 395 out of 4556 percentage completed: 8.67 %  
## Process 396 out of 4556 percentage completed: 8.69 %  
## Process 397 out of 4556 percentage completed: 8.71 %  
## Process 398 out of 4556 percentage completed: 8.74 %  
## Process 399 out of 4556 percentage completed: 8.76 %  
## Process 400 out of 4556 percentage completed: 8.78 %  
## Process 401 out of 4556 percentage completed: 8.8 %  
## Process 402 out of 4556 percentage completed: 8.82 %  
## Process 403 out of 4556 percentage completed: 8.85 %  
## Process 404 out of 4556 percentage completed: 8.87 %  
## Process 405 out of 4556 percentage completed: 8.89 %  
## Process 406 out of 4556 percentage completed: 8.91 %  
## Process 407 out of 4556 percentage completed: 8.93 %  
## Process 408 out of 4556 percentage completed: 8.96 %  
## Process 409 out of 4556 percentage completed: 8.98 %  
## Process 410 out of 4556 percentage completed: 9 %  
## Process 411 out of 4556 percentage completed: 9.02 %  
## Process 412 out of 4556 percentage completed: 9.04 %  
## Process 413 out of 4556 percentage completed: 9.06 %  
## Process 414 out of 4556 percentage completed: 9.09 %  
## Process 415 out of 4556 percentage completed: 9.11 %  
## Process 416 out of 4556 percentage completed: 9.13 %  
## Process 417 out of 4556 percentage completed: 9.15 %  
## Process 418 out of 4556 percentage completed: 9.17 %  
## Process 419 out of 4556 percentage completed: 9.2 %  
## Process 420 out of 4556 percentage completed: 9.22 %  
## Process 421 out of 4556 percentage completed: 9.24 %  
## Process 422 out of 4556 percentage completed: 9.26 %  
## Process 423 out of 4556 percentage completed: 9.28 %  
## Process 424 out of 4556 percentage completed: 9.31 %  
## Process 425 out of 4556 percentage completed: 9.33 %  
## Process 426 out of 4556 percentage completed: 9.35 %  
## Process 427 out of 4556 percentage completed: 9.37 %  
## Process 428 out of 4556 percentage completed: 9.39 %  
## Process 429 out of 4556 percentage completed: 9.42 %  
## Process 430 out of 4556 percentage completed: 9.44 %  
## Process 431 out of 4556 percentage completed: 9.46 %  
## Process 432 out of 4556 percentage completed: 9.48 %  
## Process 433 out of 4556 percentage completed: 9.5 %  
## Process 434 out of 4556 percentage completed: 9.53 %  
## Process 435 out of 4556 percentage completed: 9.55 %  
## Process 436 out of 4556 percentage completed: 9.57 %  
## Process 437 out of 4556 percentage completed: 9.59 %  
## Process 438 out of 4556 percentage completed: 9.61 %  
## Process 439 out of 4556 percentage completed: 9.64 %  
## Process 440 out of 4556 percentage completed: 9.66 %  
## Process 441 out of 4556 percentage completed: 9.68 %  
## Process 442 out of 4556 percentage completed: 9.7 %  
## Process 443 out of 4556 percentage completed: 9.72 %  
## Process 444 out of 4556 percentage completed: 9.75 %  
## Process 445 out of 4556 percentage completed: 9.77 %  
## Process 446 out of 4556 percentage completed: 9.79 %  
## Process 447 out of 4556 percentage completed: 9.81 %  
## Process 448 out of 4556 percentage completed: 9.83 %

## Process 449 out of 4556 percentage completed: 9.86 %  
## Process 450 out of 4556 percentage completed: 9.88 %  
## Process 451 out of 4556 percentage completed: 9.9 %  
## Process 452 out of 4556 percentage completed: 9.92 %  
## Process 453 out of 4556 percentage completed: 9.94 %  
## Process 454 out of 4556 percentage completed: 9.96 %  
## Process 455 out of 4556 percentage completed: 9.99 %  
## Process 456 out of 4556 percentage completed: 10.01 %  
## Process 457 out of 4556 percentage completed: 10.03 %  
## Process 458 out of 4556 percentage completed: 10.05 %  
## Process 459 out of 4556 percentage completed: 10.07 %  
## Process 460 out of 4556 percentage completed: 10.1 %  
## Process 461 out of 4556 percentage completed: 10.12 %  
## Process 462 out of 4556 percentage completed: 10.14 %  
## Process 463 out of 4556 percentage completed: 10.16 %  
## Process 464 out of 4556 percentage completed: 10.18 %  
## Process 465 out of 4556 percentage completed: 10.21 %  
## Process 466 out of 4556 percentage completed: 10.23 %  
## Process 467 out of 4556 percentage completed: 10.25 %  
## Process 468 out of 4556 percentage completed: 10.27 %  
## Process 469 out of 4556 percentage completed: 10.29 %  
## Process 470 out of 4556 percentage completed: 10.32 %  
## Process 471 out of 4556 percentage completed: 10.34 %  
## Process 472 out of 4556 percentage completed: 10.36 %  
## Process 473 out of 4556 percentage completed: 10.38 %  
## Process 474 out of 4556 percentage completed: 10.4 %  
## Process 475 out of 4556 percentage completed: 10.43 %  
## Process 476 out of 4556 percentage completed: 10.45 %  
## Process 477 out of 4556 percentage completed: 10.47 %  
## Process 478 out of 4556 percentage completed: 10.49 %  
## Process 479 out of 4556 percentage completed: 10.51 %  
## Process 480 out of 4556 percentage completed: 10.54 %  
## Process 481 out of 4556 percentage completed: 10.56 %  
## Process 482 out of 4556 percentage completed: 10.58 %  
## Process 483 out of 4556 percentage completed: 10.6 %  
## Process 484 out of 4556 percentage completed: 10.62 %  
## Process 485 out of 4556 percentage completed: 10.65 %  
## Process 486 out of 4556 percentage completed: 10.67 %  
## Process 487 out of 4556 percentage completed: 10.69 %  
## Process 488 out of 4556 percentage completed: 10.71 %  
## Process 489 out of 4556 percentage completed: 10.73 %  
## Process 490 out of 4556 percentage completed: 10.76 %  
## Process 491 out of 4556 percentage completed: 10.78 %  
## Process 492 out of 4556 percentage completed: 10.8 %  
## Process 493 out of 4556 percentage completed: 10.82 %  
## Process 494 out of 4556 percentage completed: 10.84 %  
## Process 495 out of 4556 percentage completed: 10.86 %  
## Process 496 out of 4556 percentage completed: 10.89 %  
## Process 497 out of 4556 percentage completed: 10.91 %  
## Process 498 out of 4556 percentage completed: 10.93 %  
## Process 499 out of 4556 percentage completed: 10.95 %  
## Process 500 out of 4556 percentage completed: 10.97 %  
## Process 501 out of 4556 percentage completed: 11 %  
## Process 502 out of 4556 percentage completed: 11.02 %

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

## Process 773 out of 4556 percentage completed: 16.97 %  
## Process 774 out of 4556 percentage completed: 16.99 %  
## Process 775 out of 4556 percentage completed: 17.01 %  
## Process 776 out of 4556 percentage completed: 17.03 %  
## Process 777 out of 4556 percentage completed: 17.05 %  
## Process 778 out of 4556 percentage completed: 17.08 %  
## Process 779 out of 4556 percentage completed: 17.1 %  
## Process 780 out of 4556 percentage completed: 17.12 %  
## Process 781 out of 4556 percentage completed: 17.14 %  
## Process 782 out of 4556 percentage completed: 17.16 %  
## Process 783 out of 4556 percentage completed: 17.19 %  
## Process 784 out of 4556 percentage completed: 17.21 %  
## Process 785 out of 4556 percentage completed: 17.23 %  
## Process 786 out of 4556 percentage completed: 17.25 %  
## Process 787 out of 4556 percentage completed: 17.27 %  
## Process 788 out of 4556 percentage completed: 17.3 %  
## Process 789 out of 4556 percentage completed: 17.32 %  
## Process 790 out of 4556 percentage completed: 17.34 %  
## Process 791 out of 4556 percentage completed: 17.36 %  
## Process 792 out of 4556 percentage completed: 17.38 %  
## Process 793 out of 4556 percentage completed: 17.41 %  
## Process 794 out of 4556 percentage completed: 17.43 %  
## Process 795 out of 4556 percentage completed: 17.45 %  
## Process 796 out of 4556 percentage completed: 17.47 %  
## Process 797 out of 4556 percentage completed: 17.49 %  
## Process 798 out of 4556 percentage completed: 17.52 %  
## Process 799 out of 4556 percentage completed: 17.54 %  
## Process 800 out of 4556 percentage completed: 17.56 %  
## Process 801 out of 4556 percentage completed: 17.58 %  
## Process 802 out of 4556 percentage completed: 17.6 %  
## Process 803 out of 4556 percentage completed: 17.63 %  
## Process 804 out of 4556 percentage completed: 17.65 %  
## Process 805 out of 4556 percentage completed: 17.67 %  
## Process 806 out of 4556 percentage completed: 17.69 %  
## Process 807 out of 4556 percentage completed: 17.71 %  
## Process 808 out of 4556 percentage completed: 17.73 %  
## Process 809 out of 4556 percentage completed: 17.76 %  
## Process 810 out of 4556 percentage completed: 17.78 %  
## Process 811 out of 4556 percentage completed: 17.8 %  
## Process 812 out of 4556 percentage completed: 17.82 %  
## Process 813 out of 4556 percentage completed: 17.84 %  
## Process 814 out of 4556 percentage completed: 17.87 %  
## Process 815 out of 4556 percentage completed: 17.89 %  
## Process 816 out of 4556 percentage completed: 17.91 %  
## Process 817 out of 4556 percentage completed: 17.93 %  
## Process 818 out of 4556 percentage completed: 17.95 %  
## Process 819 out of 4556 percentage completed: 17.98 %  
## Process 820 out of 4556 percentage completed: 18 %  
## Process 821 out of 4556 percentage completed: 18.02 %  
## Process 822 out of 4556 percentage completed: 18.04 %  
## Process 823 out of 4556 percentage completed: 18.06 %  
## Process 824 out of 4556 percentage completed: 18.09 %  
## Process 825 out of 4556 percentage completed: 18.11 %  
## Process 826 out of 4556 percentage completed: 18.13 %

[illegible]

[illegible]





[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```

## Process 1745 out of 4556 percentage completed: 38.3 %
## Process 1746 out of 4556 percentage completed: 38.32 %
## Process 1747 out of 4556 percentage completed: 38.35 %
## Process 1748 out of 4556 percentage completed: 38.37 %
## Process 1749 out of 4556 percentage completed: 38.39 %
## Process 1750 out of 4556 percentage completed: 38.41 %
## Process 1751 out of 4556 percentage completed: 38.43 %
## Process 1752 out of 4556 percentage completed: 38.45 %
## ERROR : cannot open the connection
## Process 1754 out of 4556 percentage completed: 38.5 %
## Process 1755 out of 4556 percentage completed: 38.52 %
## ERROR : cannot open the connection
## Process 1757 out of 4556 percentage completed: 38.56 %
## Process 1758 out of 4556 percentage completed: 38.59 %
## Process 1759 out of 4556 percentage completed: 38.61 %
## Process 1760 out of 4556 percentage completed: 38.63 %
## Process 1761 out of 4556 percentage completed: 38.65 %
## Process 1762 out of 4556 percentage completed: 38.67 %
## Process 1763 out of 4556 percentage completed: 38.7 %
## Process 1764 out of 4556 percentage completed: 38.72 %
## Process 1765 out of 4556 percentage completed: 38.74 %
## Process 1766 out of 4556 percentage completed: 38.76 %
## Process 1767 out of 4556 percentage completed: 38.78 %
## Process 1768 out of 4556 percentage completed: 38.81 %
## Process 1769 out of 4556 percentage completed: 38.83 %
## Process 1770 out of 4556 percentage completed: 38.85 %
## Process 1771 out of 4556 percentage completed: 38.87 %
## Process 1772 out of 4556 percentage completed: 38.89 %
## Process 1773 out of 4556 percentage completed: 38.92 %
## Process 1774 out of 4556 percentage completed: 38.94 %
## Process 1775 out of 4556 percentage completed: 38.96 %
## Process 1776 out of 4556 percentage completed: 38.98 %
## Process 1777 out of 4556 percentage completed: 39 %
## Process 1778 out of 4556 percentage completed: 39.03 %
## Process 1779 out of 4556 percentage completed: 39.05 %
## Process 1780 out of 4556 percentage completed: 39.07 %
## Process 1781 out of 4556 percentage completed: 39.09 %
## Process 1782 out of 4556 percentage completed: 39.11 %
## Process 1783 out of 4556 percentage completed: 39.14 %
## Process 1784 out of 4556 percentage completed: 39.16 %
## Process 1785 out of 4556 percentage completed: 39.18 %
## Process 1786 out of 4556 percentage completed: 39.2 %
## Process 1787 out of 4556 percentage completed: 39.22 %
## Process 1788 out of 4556 percentage completed: 39.24 %
## Process 1789 out of 4556 percentage completed: 39.27 %
## Process 1790 out of 4556 percentage completed: 39.29 %
## Process 1791 out of 4556 percentage completed: 39.31 %
## Process 1792 out of 4556 percentage completed: 39.33 %
## Process 1793 out of 4556 percentage completed: 39.35 %
## Process 1794 out of 4556 percentage completed: 39.38 %
## Process 1795 out of 4556 percentage completed: 39.4 %
## Process 1796 out of 4556 percentage completed: 39.42 %
## Process 1797 out of 4556 percentage completed: 39.44 %
## Process 1798 out of 4556 percentage completed: 39.46 %

```

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```
## Process 2231 out of 4556 percentage completed: 48.97 %
## ERROR : Failed to parse text
## Process 2233 out of 4556 percentage completed: 49.01 %
## Process 2234 out of 4556 percentage completed: 49.03 %
## Process 2235 out of 4556 percentage completed: 49.06 %
## Process 2236 out of 4556 percentage completed: 49.08 %
## Process 2237 out of 4556 percentage completed: 49.1 %
## Process 2238 out of 4556 percentage completed: 49.12 %
## Process 2239 out of 4556 percentage completed: 49.14 %
## Process 2240 out of 4556 percentage completed: 49.17 %
## Process 2241 out of 4556 percentage completed: 49.19 %
## Process 2242 out of 4556 percentage completed: 49.21 %
## Process 2243 out of 4556 percentage completed: 49.23 %
## Process 2244 out of 4556 percentage completed: 49.25 %
## Process 2245 out of 4556 percentage completed: 49.28 %
## Process 2246 out of 4556 percentage completed: 49.3 %
## Process 2247 out of 4556 percentage completed: 49.32 %
## Process 2248 out of 4556 percentage completed: 49.34 %
## Process 2249 out of 4556 percentage completed: 49.36 %
## Process 2250 out of 4556 percentage completed: 49.39 %
## Process 2251 out of 4556 percentage completed: 49.41 %
## Process 2252 out of 4556 percentage completed: 49.43 %
## Process 2253 out of 4556 percentage completed: 49.45 %
## Process 2254 out of 4556 percentage completed: 49.47 %
## Process 2255 out of 4556 percentage completed: 49.5 %
## Process 2256 out of 4556 percentage completed: 49.52 %
## Process 2257 out of 4556 percentage completed: 49.54 %
## Process 2258 out of 4556 percentage completed: 49.56 %
## Process 2259 out of 4556 percentage completed: 49.58 %
## Process 2260 out of 4556 percentage completed: 49.6 %
## Process 2261 out of 4556 percentage completed: 49.63 %
## Process 2262 out of 4556 percentage completed: 49.65 %
## Process 2263 out of 4556 percentage completed: 49.67 %
## Process 2264 out of 4556 percentage completed: 49.69 %
## Process 2265 out of 4556 percentage completed: 49.71 %
## Process 2266 out of 4556 percentage completed: 49.74 %
## Process 2267 out of 4556 percentage completed: 49.76 %
## Process 2268 out of 4556 percentage completed: 49.78 %
## Process 2269 out of 4556 percentage completed: 49.8 %
## Process 2270 out of 4556 percentage completed: 49.82 %
## Process 2271 out of 4556 percentage completed: 49.85 %
## Process 2272 out of 4556 percentage completed: 49.87 %
## Process 2273 out of 4556 percentage completed: 49.89 %
## Process 2274 out of 4556 percentage completed: 49.91 %
## Process 2275 out of 4556 percentage completed: 49.93 %
## Process 2276 out of 4556 percentage completed: 49.96 %
## Process 2277 out of 4556 percentage completed: 49.98 %
## Process 2278 out of 4556 percentage completed: 50 %
## Process 2279 out of 4556 percentage completed: 50.02 %
## Process 2280 out of 4556 percentage completed: 50.04 %
## Process 2281 out of 4556 percentage completed: 50.07 %
## Process 2282 out of 4556 percentage completed: 50.09 %
## Process 2283 out of 4556 percentage completed: 50.11 %
## Process 2284 out of 4556 percentage completed: 50.13 %
```



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

```

## Process 4175 out of 4556 percentage completed: 91.64 %
## Process 4176 out of 4556 percentage completed: 91.66 %
## Process 4177 out of 4556 percentage completed: 91.68 %
## Process 4178 out of 4556 percentage completed: 91.7 %
## Process 4179 out of 4556 percentage completed: 91.73 %
## Process 4180 out of 4556 percentage completed: 91.75 %
## Process 4181 out of 4556 percentage completed: 91.77 %
## Process 4182 out of 4556 percentage completed: 91.79 %
## Process 4183 out of 4556 percentage completed: 91.81 %
## Process 4184 out of 4556 percentage completed: 91.83 %
## Process 4185 out of 4556 percentage completed: 91.86 %
## Process 4186 out of 4556 percentage completed: 91.88 %
## Process 4187 out of 4556 percentage completed: 91.9 %
## Process 4188 out of 4556 percentage completed: 91.92 %
## Process 4189 out of 4556 percentage completed: 91.94 %
## Process 4190 out of 4556 percentage completed: 91.97 %
## Process 4191 out of 4556 percentage completed: 91.99 %
## Process 4192 out of 4556 percentage completed: 92.01 %
## Process 4193 out of 4556 percentage completed: 92.03 %
## Process 4194 out of 4556 percentage completed: 92.05 %
## Process 4195 out of 4556 percentage completed: 92.08 %
## Process 4196 out of 4556 percentage completed: 92.1 %
## Process 4197 out of 4556 percentage completed: 92.12 %
## Process 4198 out of 4556 percentage completed: 92.14 %
## Process 4199 out of 4556 percentage completed: 92.16 %
## Process 4200 out of 4556 percentage completed: 92.19 %
## Process 4201 out of 4556 percentage completed: 92.21 %
## Process 4202 out of 4556 percentage completed: 92.23 %
## Process 4203 out of 4556 percentage completed: 92.25 %
## Process 4204 out of 4556 percentage completed: 92.27 %
## Process 4205 out of 4556 percentage completed: 92.3 %
## Process 4206 out of 4556 percentage completed: 92.32 %
## Process 4207 out of 4556 percentage completed: 92.34 %
## Process 4208 out of 4556 percentage completed: 92.36 %
## Process 4209 out of 4556 percentage completed: 92.38 %
## Process 4210 out of 4556 percentage completed: 92.41 %
## Process 4211 out of 4556 percentage completed: 92.43 %
## Process 4212 out of 4556 percentage completed: 92.45 %
## Process 4213 out of 4556 percentage completed: 92.47 %
## Process 4214 out of 4556 percentage completed: 92.49 %
## Process 4215 out of 4556 percentage completed: 92.52 %
## Process 4216 out of 4556 percentage completed: 92.54 %
## Process 4217 out of 4556 percentage completed: 92.56 %
## Process 4218 out of 4556 percentage completed: 92.58 %
## Process 4219 out of 4556 percentage completed: 92.6 %
## ERROR : cannot open the connection
## Process 4221 out of 4556 percentage completed: 92.65 %
## Process 4222 out of 4556 percentage completed: 92.67 %
## Process 4223 out of 4556 percentage completed: 92.69 %
## Process 4224 out of 4556 percentage completed: 92.71 %
## Process 4225 out of 4556 percentage completed: 92.73 %
## Process 4226 out of 4556 percentage completed: 92.76 %
## Process 4227 out of 4556 percentage completed: 92.78 %
## Process 4228 out of 4556 percentage completed: 92.8 %

```

```

## Process 4229 out of 4556 percentage completed: 92.82 %
## Process 4230 out of 4556 percentage completed: 92.84 %
## Process 4231 out of 4556 percentage completed: 92.87 %
## Process 4232 out of 4556 percentage completed: 92.89 %
## Process 4233 out of 4556 percentage completed: 92.91 %
## Process 4234 out of 4556 percentage completed: 92.93 %
## Process 4235 out of 4556 percentage completed: 92.95 %
## Process 4236 out of 4556 percentage completed: 92.98 %
## Process 4237 out of 4556 percentage completed: 93 %
## Process 4238 out of 4556 percentage completed: 93.02 %
## Process 4239 out of 4556 percentage completed: 93.04 %
## Process 4240 out of 4556 percentage completed: 93.06 %
## Process 4241 out of 4556 percentage completed: 93.09 %
## Process 4242 out of 4556 percentage completed: 93.11 %
## Process 4243 out of 4556 percentage completed: 93.13 %
## Process 4244 out of 4556 percentage completed: 93.15 %
## Process 4245 out of 4556 percentage completed: 93.17 %
## Process 4246 out of 4556 percentage completed: 93.2 %
## Process 4247 out of 4556 percentage completed: 93.22 %
## Process 4248 out of 4556 percentage completed: 93.24 %
## Process 4249 out of 4556 percentage completed: 93.26 %
## Process 4250 out of 4556 percentage completed: 93.28 %
## ERROR : cannot open the connection
## Process 4252 out of 4556 percentage completed: 93.33 %
## Process 4253 out of 4556 percentage completed: 93.35 %
## Process 4254 out of 4556 percentage completed: 93.37 %
## Process 4255 out of 4556 percentage completed: 93.39 %
## ERROR : cannot open the connection
## Process 4257 out of 4556 percentage completed: 93.44 %
## Process 4258 out of 4556 percentage completed: 93.46 %
## Process 4259 out of 4556 percentage completed: 93.48 %
## Process 4260 out of 4556 percentage completed: 93.5 %
## Process 4261 out of 4556 percentage completed: 93.53 %
## Process 4262 out of 4556 percentage completed: 93.55 %
## Process 4263 out of 4556 percentage completed: 93.57 %
## Process 4264 out of 4556 percentage completed: 93.59 %
## Process 4265 out of 4556 percentage completed: 93.61 %
## Process 4266 out of 4556 percentage completed: 93.63 %
## Process 4267 out of 4556 percentage completed: 93.66 %
## Process 4268 out of 4556 percentage completed: 93.68 %
## Process 4269 out of 4556 percentage completed: 93.7 %
## Process 4270 out of 4556 percentage completed: 93.72 %
## Process 4271 out of 4556 percentage completed: 93.74 %
## Process 4272 out of 4556 percentage completed: 93.77 %
## Process 4273 out of 4556 percentage completed: 93.79 %
## Process 4274 out of 4556 percentage completed: 93.81 %
## Process 4275 out of 4556 percentage completed: 93.83 %
## Process 4276 out of 4556 percentage completed: 93.85 %
## Process 4277 out of 4556 percentage completed: 93.88 %
## Process 4278 out of 4556 percentage completed: 93.9 %
## Process 4279 out of 4556 percentage completed: 93.92 %
## Process 4280 out of 4556 percentage completed: 93.94 %
## Process 4281 out of 4556 percentage completed: 93.96 %
## Process 4282 out of 4556 percentage completed: 93.99 %

```

```

## Process 4283 out of 4556 percentage completed: 94.01 %
## Process 4284 out of 4556 percentage completed: 94.03 %
## Process 4285 out of 4556 percentage completed: 94.05 %
## Process 4286 out of 4556 percentage completed: 94.07 %
## Process 4287 out of 4556 percentage completed: 94.1 %
## Process 4288 out of 4556 percentage completed: 94.12 %
## Process 4289 out of 4556 percentage completed: 94.14 %
## Process 4290 out of 4556 percentage completed: 94.16 %
## Process 4291 out of 4556 percentage completed: 94.18 %
## Process 4292 out of 4556 percentage completed: 94.21 %
## Process 4293 out of 4556 percentage completed: 94.23 %
## Process 4294 out of 4556 percentage completed: 94.25 %
## ERROR : cannot open the connection
## Process 4296 out of 4556 percentage completed: 94.29 %
## Process 4297 out of 4556 percentage completed: 94.32 %
## Process 4298 out of 4556 percentage completed: 94.34 %
## Process 4299 out of 4556 percentage completed: 94.36 %
## Process 4300 out of 4556 percentage completed: 94.38 %
## Process 4301 out of 4556 percentage completed: 94.4 %
## Process 4302 out of 4556 percentage completed: 94.42 %
## Process 4303 out of 4556 percentage completed: 94.45 %
## Process 4304 out of 4556 percentage completed: 94.47 %
## Process 4305 out of 4556 percentage completed: 94.49 %
## Process 4306 out of 4556 percentage completed: 94.51 %
## Process 4307 out of 4556 percentage completed: 94.53 %
## Process 4308 out of 4556 percentage completed: 94.56 %
## Process 4309 out of 4556 percentage completed: 94.58 %
## Process 4310 out of 4556 percentage completed: 94.6 %
## Process 4311 out of 4556 percentage completed: 94.62 %
## Process 4312 out of 4556 percentage completed: 94.64 %
## Process 4313 out of 4556 percentage completed: 94.67 %
## Process 4314 out of 4556 percentage completed: 94.69 %
## Process 4315 out of 4556 percentage completed: 94.71 %
## Process 4316 out of 4556 percentage completed: 94.73 %
## Process 4317 out of 4556 percentage completed: 94.75 %
## Process 4318 out of 4556 percentage completed: 94.78 %
## Process 4319 out of 4556 percentage completed: 94.8 %
## Process 4320 out of 4556 percentage completed: 94.82 %
## Process 4321 out of 4556 percentage completed: 94.84 %
## Process 4322 out of 4556 percentage completed: 94.86 %
## Process 4323 out of 4556 percentage completed: 94.89 %
## Process 4324 out of 4556 percentage completed: 94.91 %
## Process 4325 out of 4556 percentage completed: 94.93 %
## Process 4326 out of 4556 percentage completed: 94.95 %
## Process 4327 out of 4556 percentage completed: 94.97 %
## Process 4328 out of 4556 percentage completed: 95 %
## Process 4329 out of 4556 percentage completed: 95.02 %
## Process 4330 out of 4556 percentage completed: 95.04 %
## Process 4331 out of 4556 percentage completed: 95.06 %
## Process 4332 out of 4556 percentage completed: 95.08 %
## Process 4333 out of 4556 percentage completed: 95.11 %
## Process 4334 out of 4556 percentage completed: 95.13 %
## Process 4335 out of 4556 percentage completed: 95.15 %
## Process 4336 out of 4556 percentage completed: 95.17 %

```

[illegible]

[illegible]



```
## Process 4445 out of 4556 percentage completed: 97.56 %
## Process 4446 out of 4556 percentage completed: 97.59 %
## Process 4447 out of 4556 percentage completed: 97.61 %
## Process 4448 out of 4556 percentage completed: 97.63 %
## Process 4449 out of 4556 percentage completed: 97.65 %
## Process 4450 out of 4556 percentage completed: 97.67 %
## Process 4451 out of 4556 percentage completed: 97.7 %
## Process 4452 out of 4556 percentage completed: 97.72 %
## Process 4453 out of 4556 percentage completed: 97.74 %
## Process 4454 out of 4556 percentage completed: 97.76 %
## Process 4455 out of 4556 percentage completed: 97.78 %
## Process 4456 out of 4556 percentage completed: 97.81 %
## Process 4457 out of 4556 percentage completed: 97.83 %
## Process 4458 out of 4556 percentage completed: 97.85 %
## ERROR : cannot open the connection
## Process 4460 out of 4556 percentage completed: 97.89 %
## Process 4461 out of 4556 percentage completed: 97.91 %
## Process 4462 out of 4556 percentage completed: 97.94 %
## Process 4463 out of 4556 percentage completed: 97.96 %
## Process 4464 out of 4556 percentage completed: 97.98 %
## Process 4465 out of 4556 percentage completed: 98 %
## Process 4466 out of 4556 percentage completed: 98.02 %
## Process 4467 out of 4556 percentage completed: 98.05 %
## Process 4468 out of 4556 percentage completed: 98.07 %
## Process 4469 out of 4556 percentage completed: 98.09 %
## Process 4470 out of 4556 percentage completed: 98.11 %
## Process 4471 out of 4556 percentage completed: 98.13 %
## Process 4472 out of 4556 percentage completed: 98.16 %
## Process 4473 out of 4556 percentage completed: 98.18 %
## Process 4474 out of 4556 percentage completed: 98.2 %
## Process 4475 out of 4556 percentage completed: 98.22 %
## Process 4476 out of 4556 percentage completed: 98.24 %
## Process 4477 out of 4556 percentage completed: 98.27 %
## Process 4478 out of 4556 percentage completed: 98.29 %
## Process 4479 out of 4556 percentage completed: 98.31 %
## Process 4480 out of 4556 percentage completed: 98.33 %
## Process 4481 out of 4556 percentage completed: 98.35 %
## Process 4482 out of 4556 percentage completed: 98.38 %
## Process 4483 out of 4556 percentage completed: 98.4 %
## Process 4484 out of 4556 percentage completed: 98.42 %
## Process 4485 out of 4556 percentage completed: 98.44 %
## Process 4486 out of 4556 percentage completed: 98.46 %
## Process 4487 out of 4556 percentage completed: 98.49 %
## Process 4488 out of 4556 percentage completed: 98.51 %
## Process 4489 out of 4556 percentage completed: 98.53 %
## Process 4490 out of 4556 percentage completed: 98.55 %
## Process 4491 out of 4556 percentage completed: 98.57 %
## Process 4492 out of 4556 percentage completed: 98.6 %
## Process 4493 out of 4556 percentage completed: 98.62 %
## Process 4494 out of 4556 percentage completed: 98.64 %
## Process 4495 out of 4556 percentage completed: 98.66 %
## Process 4496 out of 4556 percentage completed: 98.68 %
## Process 4497 out of 4556 percentage completed: 98.71 %
## Process 4498 out of 4556 percentage completed: 98.73 %
```

[illegible]

```
## Process 4553 out of 4556 percentage completed: 99.93 %
## Process 4554 out of 4556 percentage completed: 99.96 %
## Process 4555 out of 4556 percentage completed: 99.98 %
## Process 4556 out of 4556 percentage completed: 100 %
```

```
head(Book_total)
```

```
##           Book           Title
## 1 紅樓夢(曹雪芹) 第一回 甄士隱夢幻識通靈 賈雨村風塵懷閨秀
## 2 紅樓夢(曹雪芹) 第二回 賈夫人仙逝揚州城 冷子興演①榮國府
## 3 紅樓夢(曹雪芹)           第三回
## 4 紅樓夢(曹雪芹)           第四回
## 5 紅樓夢(曹雪芹)           第五回
## 6 紅樓夢(曹雪芹)           第六回
##
## 1
```

### 3. Save the data

#### 3.1 Save as an Excel file

```
write.xlsx(Book_total, "ClassicChineseLiterature_CorpusCourse.xlsx")
```

```
## Warning in wb$writeData(df = x, colNames = colNames, sheet = sheet, startCol = startCol, : 朱四府道：
## 網絡圖書獨家推出
##
## is truncated.
## Number of characters exceed the limit of 32767.

## Warning in wb$writeData(df = x, colNames = colNames, sheet = sheet, startCol = startCol, : , 探其巢穴
## 網絡圖書掃描校對
##
## is truncated.
## Number of characters exceed the limit of 32767.

## Warning in wb$writeData(df = x, colNames = colNames, sheet = sheet, startCol = startCol, : 雪。到了冬
## 中文東西網 整理
```

```
##  
## is truncated.  
## Number of characters exceed the limit of 32767.
```

### 3.2 Save as an RData file

```
save(Book_total, file = "ClassicChineseLiterature_CorpusCourse.Rdata")
```