

Makes one table from all chapters

Casey Crary

2024-04-02

This file takes all the data from “data-raw-depaginate” (which will be changed to “data-raw-dehyphenate”), and appends all the txt files into one. Aspirational goal: make this output an rds file instead.

```
write("", "../data/output.txt")          # create file to output to

# appends a given chapter_name onto output.txt
append_chapter <- function(chapter_name) {
  to_write_to <- "../data/output.txt"
  file_name <- paste0("../data-raw-depaginate/", chapter_name, ".txt")
  chap_data <- readLines(file_name)

  write(chap_data, file=to_write_to, append=TRUE)
  write("\n", file=to_write_to, append = TRUE)

  return(NULL)
}

# does so for all chapters (NOT appendix, needs to be added somehow)
for(i in 0:28) {
  chapter_num <- sprintf("%02d", i)
  chapter_name <- paste("chapter", chapter_num, "_cleaned", sep="")

  append_chapter(chapter_name)
}

# appendix

# Once appendix cleaned exists this should work? XX
# append_chapter("appendix_cleaned.txt")
```

```

# proof we can then put output.txt into an R object
HistoryAmherstCollege <- readLines("../data/output.txt")

# makes tibble with line number and chapter number
hac_tibble <- HistoryAmherstCollege |> tibble()

hac_tibble_mod <- hac_tibble |>
  rename(text = HistoryAmherstCollege) |>
  mutate(linenumber = row_number(),
         chapter = cumsum(
           str_detect(text,
                      regex("^CHAPTER [\\DIVXLC]+",
                           ignore_case = FALSE))
         )
  )
hac_tibble_mod

```

```
# A tibble: 27,445 x 3
```

	text <chr>	linenumber <int>	chapter <int>
1	" "	1	0
2	"Amherst College DUring ITS FIRST HALF CENTURY. 1821-187~	2	0
3	"BY W. S. TYLER, "	3	0
4	"OF THE CLASS OF 1830, "	4	0
5	"Williston Professor of the Greek Language and Literature~	5	0
6	" "	6	0
7	"SPRINGFIELD, MASS.: "	7	0
8	"CLARK W. BRYAN AND COMPANY. "	8	0
9	"1873. "	9	0
10	" "	10	0

```
# i 27,435 more rows
```

```
# usethis::use_data(HistoryAmherstCollege, overwrite = TRUE)
```