

# Wrangling Page Breaks & Subtitles

Adam, Frankie

2024-04-04

## Last Run:

2024-04-04 15:34:32

```
fix_up_lines <- function(file) {  
  file <- chapter_name  
  chapter_lines <- readLines(file)  
  n <- length(chapter_lines)  
  for (i in 1 : n) {  
    chapter_lines[i] <- str_trim(chapter_lines[i])  
    if (str_detect(chapter_lines[i], "-$")) {  
      next_line <- 1  
      if (str_trim(chapter_lines[i+1]) == "") next_line <- 2  
      padding <- str_match(chapter_lines[i+next_line], "[\\w+\\d]+[:punct:]?")  
      chapter_lines[i] <- paste(str_sub(chapter_lines[i], start=1, end=-2), padding, sep="")  
      chapter_lines[i] <- str_trim(chapter_lines[i])  
      chapter_lines[i+next_line] <- str_replace(chapter_lines[i+next_line], fixed(padding), "")  
    }  
  }  
  return(chapter_lines)  
}
```

```
path <- getwd()  
for(i in 0:29) {  
  chapter_num <- sprintf("%02d", i)  
  chapter_name <- paste(path,  
    "../data-raw-depaginate/chapter",  
    chapter_num,  
    "_cleaned.txt", sep = "")  
}
```

```
table <- fix_up_lines(chapter_name)
write.table(table, paste0(path,
                          "../data-raw-dehyphenate/chapter",
                          chapter_num,
                          "_dehyphenate.txt", sep = ""),
            quote = FALSE, row.names = FALSE, col.names = FALSE)
}
```