

R.scripts

NICER

chatToText

```
#chatToText
#2020-01-17 sugiura@nagoya-u.jp
#CHAT フォーマットのファイルから、
#本文のメインティアのテキストだけを抜き出して
#もとのファイル名に .data を付け足したファイル名で保存する。

chatToText <- function(){
  # ディレクトリー内のすべてのテキストファイルを対象に
  here <- getwd()
  files <- list.files(here, pattern="\\.txt$")
  # 読み込むファイル名で、.txt という拡張子でファイル名が終わるものを指定
  for (i in files){

    lines.tmp <- scan(i, what="char", sep="\n")

    data.tmp <- grep("\\*(JPN|NS)...:\\t", lines.tmp, value=T)

    body.tmp <- gsub("\\*(JPN|NS)...:\\t", "", data.tmp)

    body.tmp <- body.tmp[body.tmp != ""]

    filename <- i
    filename <- as.factor(filename)
    filename <- paste(filename, ".data", sep="")
    # もとのファイル名に .data という文字列を追加
    # ファイル名の終わりが .txt ではなくなるので再帰的に読み込まれない
    write(body.tmp, file=filename)

  }
}
```

chatToText

- ・ディレクトリーを作って、その中に、
- ・txt ファイルとして結果を保存

```
#chatToNewText
#2021-12-06 sugiura@nagoya-u.jp
#CHAT フォーマットのファイルから、
#本文のメインティアのテキストだけを抜き出して
#もとのファイル名に .new.txt を付け足したファイル名で、new というディレクトリー内に保存する。

chatToNewText <- function(){
  # ディレクトリー内のすべてのテキストファイルを対象に
  here <- getwd()
  files <- list.files(here, pattern="\\.txt$")
  # 読み込むファイル名で、.txt という拡張子でファイル名が終わるものを指定
  dir.create("text_only")
  # text_only というディレクトリーの作成

  for (i in files){

    lines.tmp <- scan(i, what="char", sep="\n")

    data.tmp <- grep("\\*(JPN|NS)...:\\t", lines.tmp, value=T)

    body.tmp <- gsub("\\*(JPN|NS)...:\\t", "", data.tmp)
```

```
body.tmp <- body.tmp[body.tmp != ""]  
filename <- i  
filename <- as.factor(filename)  
filename <- paste("text_only/", filename, ".new.txt", sep="")  
write(body.tmp, file=filename)  
}  
}
```