

R

R.scripts

myWordList.R

カレント・ディレクトリー内のすべてのテキストファイルから語彙リスト作成

```
myWordList <- function() {  
  token.all <- ""  
  files <- list.files()  
  for (i in files) {  
    token <- ""  
    lines.tmp <- scan(i, what="char")  
    body.lower <- tolower(lines.tmp)  
    body.token <- unlist(strsplit(body.lower, "¥¥¥W"))  
    body.token <- body.token[body.token != ""]  
    token.all <- c(token.all, body.token)  
  }  
  word.list <- sort(token.all)  
  word.list <- word.list[word.list != ""]  
  as.data.frame(word.list)  
}
```

- ・作業ディレクトリーを目的のディレクトリーに移動していること
 - ・ getwd() で確認
- ・テキストファイルだけが入っていることを確認
 - ・ list.files() で確認
- ・標準出力（コンソール画面）に出力される
- ・オブジェクトに保存すればそのまま data.frame として使用可能

```
sample.df <- myWordList()
```

- ・これを table() に入れば、語彙頻度一覧の完成
 - ・語彙頻度一覧をオブジェクト（データフレーム）として利用もできる

```
as.data.frame(table(sample.df))
```

as.data.frame(table(myWordList())) と入れ子式に一行で実行してもよい