

R

R.sample : 簡単な処理のサンプル

フォルダー内のデータの基本的な言語特徴を算出する

ファイル名	Criterionスコア	トピック	type	token	TTR	GI	平均単語長	平均文
-------	--------------	------	------	-------	-----	----	-------	-----

- ・ NICER 分析用
- ・ 作業ディレクトリーを分析対象のフォルダーに設定しておく
- ・ 結果は、そのフォルダーの上のフォルダーに LingFeatures.txt として保存される。

```
myLing7 <- function() {  
  files <- list.files()  
  for (i in files) {  
    lines.tmp <- scan(i, what="char", sep="\n")  
    # ファイルを選択。  
    data.tmp <- grep("¥¥*(JPN|NS)...:¥t", lines.tmp, value=T)  
    # * で始まる行のみ。  
    body.tmp <- gsub("¥¥*(JPN|NS)...:¥t", "", data.tmp)  
    # 行頭の記号とタグを削除。  
    body.tmp <- body.tmp[body.tmp != ""]  
    # 空の要素を削除する「イディオム」  
    body.lower <- tolower(body.tmp)  
    body.token <- unlist(strsplit(body.lower, "¥¥W"))  
    body.token <- body.token[body.token != ""]  
    # 空の要素を削除する「イディオム」  
    type <- length(unique(body.token))  
    token <- length(body.token)  
    TTR <- type/token  
    GI <- type/sqrt(token)  
    # ファイル名の取得は、i に入っている  
    # トピックの取得  
    topic.line <- grep("^@Topic:¥t", lines.tmp, value=T)  
    topic <- gsub("^@Topic:¥t", "", topic.line)  
    # スコアの取得  
    criterion.line <- grep("^@Criterion:¥t", lines.tmp, value=T)  
    score <- gsub("^@Criterion:¥t", "", criterion.line)  
    # Average Word Length  
    awl <- nchar(paste(body.token, collapse=""))/length(body.token)  
    # Average Sentence Length  
    asl <- length(body.token)/length(body.tmp)  
  
    cat(i, score, topic, type, token, TTR, GI, awl, asl, "\n", file="../LingFeatures.txt", append=T)  
  }  
}
```

特定の分析コードの内容の一覧を取得し、ファイルに保存する。

- ・ 次の分析コード欄を対象に、その内容を取り出す。

%Sub:GF:

- ・ 最後のコロンの後ろにタブ記号が入っている。

1. 上のスクリプトを R に読み込む
 2. 出力結果を適当なベクトルに保存する (例: result)
- ・ 結果の入ったベクトル result をファイル (例: result.txt) に書き出す。

```
> write(result, "result.txt")
```

特定の分析コードの内容の一覧を取得し、ファイルに保存する。(フォルダー内のすべてのファイルを対象に)

■ 作業前の前提

1. 処理対象のファイルを一つのフォルダーにまとめる
2. 作業ディレクトリーをそのフォルダーに設定する

■ 作業手順

1. プログラムを実行し、結果をたとえば result という変数に保存して
2. それをファイルに書き出す。(そのまま書き出すと作業ディレクトリーのフォルダー内に保存される)

```
result <- myGrepCodeFolder("¥¥*JPN...:¥¥t")
write(result, "result.txt")
```

■ 検索例

・ データ本文

```
"¥¥*JPN...:¥¥t"
```

・ 母語話者による書き換え文

```
"%NTV:¥¥t"
```

・ 母語話者によるコメント

```
"%COM:¥¥t"
```

・ 分析コード

```
"%COD:¥¥t"
```

特定の分析コードの内容の一覧を取得し、ファイルに保存する。(フォルダー内のすべてのファイルを対象に) 行頭記号をそのままにする場合

■ データ本文と添削文の両方を抽出

- ・ 正規表現 を使う。(丸カッコの中、縦棒の前後のいずれかという意味)
- ・ 行頭記号が残っていれば、どのファイルのものがわかる

```
myGrepCodeFolder2("(¥¥*JPN...:¥¥t|%NTV:¥¥t)")
```

- ・ 結果を保存する場合、一つ上のディレクトリー(..)に保存すれば、データと結果が混ざらない

- 検索結果の行頭にファイル名を表示