

R

R.scripts R のちょっとしたスクリプト

検索

- ・ grep は、検索文字列が含まれる要素を返す。
- ・ 文字列が「単語」で、要素が「文」の場合、
 - ・ その「文」の中に「単語」が含まれているかどうかを探す。
 - ・ 含まれている「文」を返す。
 - ・ 注意点：「文」の中に「単語」が2回使われている場合
 - ・ grep したらその「文」を返すだけ。その「文」にはその「単語」が含まれている
 - ・ 「文」の中に「単語」が2回含まれていることは答えてくれない。

■ myGrep

- ・ ウィンドウが開いて検索対象ファイルをえらぶ（例： myGrep("had some"）

■ myGrep2

- ・ ファイル名も指定する（例： myGrep2("had some", "JPN005.txt"）
 - ・ ただし、作業ディレクトリーが検索対象ファイルのあるディレクトリー（フォルダ）にあること

■ myGrep3.R

- ・ ディレクトリー内のすべてのファイルに対し、対象とする文字列が何回出現するかの一覧表を出力

■ myGrep4.R

Type（異なり語）の一覧の出力

■ myAllTypes.R

- ・ NICE 形式のファイルの異なり語の一覧の出力
- ・ 作業ディレクトリーをフォルダー内に
- ・ 結果を保存するファイルをフォルダー外に作成

Token（延べ語）の一覧の出力

■

- ・ NICE 形式のファイルの延べ語の一覧の出力
- ・ 作業ディレクトリーをフォルダー内に
- ・ 結果を保存するファイルをフォルダー外に作成

語彙リストの作成

■ テキストファイルを選んで作成

■ WordFreqList.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作成

- ・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()
```

■

単語リストのチェック

■ checkWordList.R

- ・Word リストにあるかどうかチェック
- ・ないものだけを出力

言語特徴量を算出

■ myTextIndex.R

- ・ディレクトリー内のすべてのテキストファイルの基礎的言語特徴量を算出

■ myIndexNICER3

- ・ファイル名・スコア・言語指標を取り出す