

R

R.scripts

grepExtract

- Stanford Parser の解析結果を対象に検索をして、
- 必要な統語関係のフレーズを抽出する。
- stringr パッケージがインストールしてあること。
- 統語解析してあるテキストファイルの入っているディレクトリーを Working Directory に設定。
- list.files() でファイルを確認。

```
grepExtract <- function(a){  
  # stringr パッケージ利用  
  #copyleft 2020-01-18 sugiura@nagoya-u.jp  
  
  hit.all <- ""  
  
  files <- list.files()  
  for (i in files) {  
    lines.tmp <- scan(i, what="char", sep="\n")  
    hit <- str_extract(lines.tmp, a)  
    hit.all <- c(hit.all, hit)  
  }  
  hit.all  
}
```

■例： 形容詞＋名詞からなる名詞句の抽出

```
grepExtract("\\(NP \\(JJ \\w+?\\) \\(NN \\w*?\\)\\)\\)")
```