
인문학 텍스트 마이닝

Text Analysis utilizing KoNLP

Text Analysis utilizing KoNLP

- 패키지 설치

```
> install.packages("KoNLP")
```

```
URL 'https://cran.rstudio.com/bin/macosx/mavericks/contrib/3.2/KoNLP_0.76.9.tgz'을 시도합니다  
Content type 'application/x-gzip' length 467717 bytes (456 KB)
```

```
=====
downloaded 456 KB
```

The downloaded binary packages are in

```
/var/folders/28/g8cf_pvx46s5phqgwr6qq7jw0000gn/T//RtmpIPv2e4/downloaded_packages
```

- 라이브러리 불러오기

```
> library(KoNLP)
```

```
필요한 패키지를 로딩중입니다: rJava
```

```
필요한 패키지를 로딩중입니다: stringr
```

```
필요한 패키지를 로딩중입니다: hash
```

```
hash-2.2.6 provided by Decision Patterns
```

```
필요한 패키지를 로딩중입니다: tau
```

```
필요한 패키지를 로딩중입니다: Sejong
```

```
Successfully Loaded Sejong Package.
```

```
Checking user defined dictionary!
```

```
다음의 패키지를 부착합니다: 'KoNLP'
```

```
The following object is masked from 'package:tau':
```

```
is.ascii
```

Text Analysis utilizing KoNLP

- extractNoun을 사용한 명사 분석(1)

```
> extractNoun("롯데마트가 판매하고 있는 흑마늘 양념 치킨이 논란이 되고 있다.")
```

```
[1] "롯데마트가" "판매" "흑마늘" "양념" "치킨" "논란"
```

- extractNoun을 사용한 명사 분석(2)

```
> sapply(c("R은 free 소프트웨어이고, [완전하게 무보증]입니다.", "일정한 조건에 따르면, 자유롭게 이것을 재배포할수가 있습니다."), extractNoun)
```

```
$`R은 free 소프트웨어이고, [완전하게 무보증]입니다.`
```

```
[1] "R" "free" "소프트웨어" "완전" "하게" "무보"
```

```
[7] "증"
```

```
$`일정한 조건에 따르면, 자유롭게 이것을 재배포할수가 있습니다.`
```

```
[1] "일정" "한" "조건" "자유" "이것"
```

```
[6] "재배포할수가"
```

Text Analysis utilizing KoNLP

- SimplePos09를 사용한 형태소 분석

```
> pos <- SimplePos09("롯데마트가 판매하고 있는 흑마늘 양념 치킨이 논란이 되고 있다.")
```

```
> pos
```

```
$롯데마트가
```

```
[1] "롯데마트가/N"
```

```
$판매하고
```

```
[1] "판매/N+하고/J"
```

```
$있는
```

```
[1] "있/P+는/E"
```

```
$흑마늘
```

```
[1] "흑마늘/N"
```

```
$양념
```

```
[1] "양념/N"
```

```
$치킨이
```

```
[1] "치킨/N+이/J"
```

```
$논란이
```

```
[1] "논란/N+이/J"
```

```
$되고
```

```
[1] "되/P+고/E"
```

```
$있다
```

```
[1] "있/P+다/E"
```

```
$.
```

```
[1] ". /S"
```

Text Analysis utilizing KoNLP

- 한글을 자음 모음으로 나타내기

```
> convertHangulStringToJamos("R는 많은 공헌자에의한 공동 프로젝트입니다")
```

```
[1] "R"      "ㄴ_ㄴ" " "      "ㅇㅏㅑ" "ㅇ_ㄴ" " "      "ㄱㅓㅇ" "ㅎㅓㄴ" "ㅈㅓ"
[10] "ㅇㅔ"    "ㅇ_ㅓ" "ㅎㅓㄴ" " "      "ㄱㅓㅇ" "ㄷㅓㅇ" " "      "표_ㅓ" "ㄹㅓ"
[19] "ㅈㅔㄱ" "ㅌ_ㅓ" "ㅇㅣㅑ" "ㄴㅣ" "ㄷㅓ"
```

- 한글을 키보드 획으로 나타내기

```
> convertHangulStringToKeyStrokes("R는 많은 공헌자에의한 공동 프로젝트입니다")
```

```
[1] "R"      "sms"  " "      "aks g" "dms"  " "      "rhd"
[8] "gjs"   "wk"   "dp"    "dml"  "gks"  " "      "rhd"
[15] "ehd"   " "    "vm"    "fh"   "wpr"  "xm"    "dlq"
[22] "s l"   "ek"
```

- 자음모음 분류를 다시 한글로 전환하기(1)

```
> str <- convertHangulStringToJamos("배포 조건의 상세한것에 대해서는 'license()' 또는 'licence()' 라고  
입력해주세요")
```

```
>
```

```
> str
```

```
[1] "ㅑㅑ"  "표ㅓ"  " "      "ㅈㅓ"  "ㄱㅓㄴ" "ㅇ_ㅓ"  " "      "ㅈㅓㅇ" "ㅈㅔ"
[10] "ㅎㅓㄴ" "ㄱㅓㅓ" "ㅇㅔ"   " "      "ㄷㅑ"  "ㅎㅑ"  "ㅈㅓ"  "ㄴ_ㄴ" " "
[19] "ㅣ"    "ㅣ"    "ㅣ"    "c"     "e"     "n"     "s"     "e"     "(
[28] ")"    "ㅣ"    " "    "ㅌㅓ"  "ㄴ_ㄴ" " "    "ㅣ"    "ㅣ"    "ㅣ"
[37] "c"    "e"    "n"    "c"     "e"     "(      ")"    "ㅣ"    " "
[46] "ㄹㅓ"  "ㄱㅓ"  " "      "ㅇㅣㅑ" "ㄹㅔㄱ" "ㅎㅑ"  "ㅈㅓ"  "ㅈㅣㅑ" "ㅈㅣ"
[55] "ㅇㅓ"
```

Text Analysis utilizing KoNLP

- 자음모음 분류를 다시 한글로 전환하기(2)

```
> str2 <-paste(str, collapse="")
```

```
>
```

```
> str2
```

```
[1] "배포 조건을 상세한것에 대해서는 'license()' 또는 'licence()' 라고  
    하고 입력해주십시오"
```

```
>
```

- 자음모음 분류를 다시 한글로 전환하기(3)

```
> HangulAutomata(str2)
```

```
[1] "배포 조건의 상세한것에 대해서는 'license()' 또는 'licence()' 라고 입력해주십시오"
```

Text Analysis utilizing KoNLP

- 시스템 사전을 사용한 형태소 분석

```
> useSystemDic()
```

```
Backup was just finished!
```

```
13 words were added to dic_user.txt.
```

```
>
```

```
> extractNoun("성긴털제비꽃은 너무 예쁘다.")
```

```
[1] "성긴털제비꽃은"
```

```
>
```

```
> SimplePos22("성긴털제비꽃은 너무 예쁘다.")
```

```
$성긴털제비꽃은
```

```
[1] "성긴털제비꽃은/NC"
```

```
$너무
```

```
[1] "너무/MA"
```

```
$예쁘다
```

```
[1] "예쁘/PA+다/EF"
```

```
$.
```

```
[1] "./SF"
```


Text Analysis utilizing KoNLP

- 세종 사전을 사용한 형태소 분석

```
> useSejongDic()
```

```
Backup was just finished!
```

```
87007 words were added to dic_user.txt.
```

```
>
```

```
> extractNoun("성긴털제비꽃은 너무 예쁘다.")
```

```
[1] "성긴털제비꽃"
```

```
>
```

```
> SimplePos22("성긴털제비꽃은 너무 예쁘다.")
```

```
$성긴털제비꽃은
```

```
[1] "성긴털제비꽃/NC+은/JX"
```

```
$너무
```

```
[1] "너무/MA"
```

```
$예쁘다
```

```
[1] "예쁘/PA+다/EF"
```

```
$.
```

```
[1] "./SF"
```

Text Analysis utilizing KoNLP

- 사용자 사전을 활용한 형태소 분석(1)

```
> useSystemDic()
```

```
Backup was just finished!
```

```
13 words were added to dic_user.txt.
```

```
>
```

```
> extractNoun("삼성전자에서는 아이폰에 대항할 무기를 준비하고 있다.")
```

```
[1] "삼성"      "전자"      "아이폰에" "대항"      "무기"      "준비"
```

- 사용자 사전을 활용한 형태소 분석(2)

```
> mergeUserDic(data.frame("삼성전자", "ncn"))
```

```
1 words were added to dic_user.txt.
```

```
>
```

```
> mergeUserDic(data.frame("아이폰", "ncn"))
```

```
1 words were added to dic_user.txt.
```

- 사용자 사전을 활용한 형태소 분석(3)

```
> extractNoun("삼성전자에서는 아이폰에 대항할 무기를 준비하고 있다.")
```

```
[1] "삼성전자" "아이폰"    "대항"      "무기"      "준비"
```