

텍스트 분석을 위한 R 패키지 설치 가이드

<https://mrchypark.github.io/textR/installation>

[pdf버전] [문의하기] [의견 및 오류 신고]

스타누르기는 콘텐츠 제작자를 춤추게 합니다.

박찬엽

2018년 09월 13일

windows 10 사용자

설치를 위한 패키지 설치

아래 코드를 R 콘솔에서 실행하세요.

```
# install_github 기능을 제공하는 remotes 패키지를 설치합니다.
install.packages("remotes")
library(remotes)
# stringi 패키지를 설치합니다.
# rtools 문제로 미리 컴파일된 버전을 설치합니다.
install.packages(
  "https://github.com/gagolews/stringi/raw/master/devel/winbuilder/3.5/stringi_1.2.4.zip",
  repos = NULL,
)
# 여러 프로그램의 설치를 지원하는 installr 패키지를 설치합니다.
install.packages("installr")
library(installr)

# 설치 화면이 나오는데, 다른 설정 변경 없이 `다음` 클릭하며 설치해주세요.
install.rtools()
# 환경 변수를 등록합니다.
Sys.setenv(PATH = paste(Sys.getenv("PATH"), "C:/Rtools/mingw_64/bin/", sep = ";"))
Sys.setenv(BINPREP = "C:/Rtools/mingw_64/bin/")
Sys.setenv(TAR = "C:/Rtools/bin/tar.exe")
```

KoNLP 패키지 설치 및 테스트

KoNLP는 가장 유명한 형태소 분석기입니다.

```
# java 인스톨을 지원하는 버전을 설치합니다.  
install_github("talgalili/installr")  
  
## 설치 이후에 Ctrl + Shift + F10 을 눌러 R session을 재시작해 주세요.  
  
library(installr)  
  
# 자바를 설치합니다.  
install.java()  
# KoNLP 패키지를 설치합니다.  
install.packages("KoNLP")  
library(KoNLP)  
  
test <- "한글 테스트 입니다."  
extractNoun(test)
```

맨 마지막 코드로 아래 결과가 나와야 합니다.

```
[1] "한글"    "테스트"
```

KoSpacing 패키지 설치 및 테스트

KoSpacing 패키지는 자동 띄어쓰기 기능을 제공합니다.

```
library(remotes)
library(installr)

# 설치 화면이 나오는데, 다른 설정 변경 없이 `다음` 클릭하며 설치해주세요.
install.conda()
install_github('mrchypark/KoSpacing')
library(KoSpacing)
set_env()
spacing("김형호영화시장분석가는 '1987'의네이버영화정보네티즌10점평에서언급된단어들을지난해12월27일부터올해1월10일")
```

아래 결과가 나오면 성공입니다.

```
[1] "김형호 영화시장 분석가는 '1987'의 네이버 영화 정보 네티즌 10점 평에서 언급된 단어들을 지난해 12월 27일부터
```

RcppMecab 패키지 설치 및 테스트

RcppMecab은 빠른 형태소 분석기입니다.

```
library(remotes)
# 설치에 필요한 환경 변수를 등록합니다.
Sys.setenv(
  PATH = paste(Sys.getenv("PATH"),
    "C:/Rtools/mingw_32/bin",
    "C:/Rtools/mingw_64/bin",
    sep = ";")
)
install_github("mrchypark/RcppMecab")
library(RcppMeCab)
# mecab의 윈도우 버전을 설치합니다.
installMecabKoForWindows()
# mecab 사전의 윈도우 버전을 설치합니다.
installMecabKoDicForWindows()
```

```
test <- "한글 테스트 입니다."
pos(test)
```

아래 결과가 나와야 합니다.

```
$`한글 테스트 입니다.`
[1] "한글/NNG"      "테스트/NNG"    "입니다/VCP"
<----->
```

글자가 깨진다면 아래 코드로 시도해 보세요.

```
test <- "한글 테스트 입니다."
# iconv 함수는 인코딩을 변경하는 함수입니다.
test <- iconv(test, to = "UTF-8")
pos(test)
```

추가 패키지 설치

tidytext 패키지는 텍스트를 tidy하게 다룰 수 있게 해줍니다.

presidentSpeechKr 패키지는 역대 대통령 연설문 텍스트를 사용할 수 있게 해주는 패키지입니다.

```
install.packages("tidytext")  
install_github("forkonlp/presidentSpeechKr")  
  
library(tidytext)  
library(presidentSpeechKr)
```

위 패키지들은 패키지를 불러오는 것이 동작하는 것을 확인해주세요.

macOS 사용자

터미널에서 설치해야 하는 도구들

1. JDK8

동영상 가이드를 참고하여 설치해주세요.

<https://www.youtube.com/watch?v=v8xZWbIASc0>

```
java -version
```

버전이 1.8.XXXXX로 표시되면 잘 설치된 것입니다.

터미널에서 설치해야 하는 도구들

2. conda3

맥의 터미널에서 아래 명령어를 수행합니다.

설치가 시작되면, 모두 yes로 입력하면서 진행하면 설치가 완료됩니다.

```
wget https://repo.continuum.io/miniconda/Miniconda3-latest-MacOSX-x86_64.sh  
bash Miniconda3-latest-MacOSX-x86_64.sh
```

터미널에서 설치해야 하는 도구들

3. mecab-ko

형태소 분석기를 설치합니다.

공식 문서를 참고하세요.

<https://bitbucket.org/eunjeon/mecab-ko-dic>

맥의 터미널에서 아래 명령어를 수행합니다.

```
wget https://bitbucket.org/eunjeon/mecab-ko/downloads/mecab-0.996-ko-0.9.2.tar.gz
tar zxfv mecab-0.996-ko-0.9.2.tar.gz
cd mecab-0.996-ko-0.9.2
./configure
make
make check
# 아래 실행 후 컴퓨터 비밀번호 입력
sudo make install
```

터미널에서 설치해야 하는 도구들

4. mecab-ko-dic

형태소 분석기의 사전을 설치합니다.

공식 문서를 참고하세요.

<https://bitbucket.org/eunjeon/mecab-ko-dic>

맥의 터미널에서 아래 명령어를 수행합니다.

```
wget https://bitbucket.org/eunjeon/mecab-ko-dic/downloads/mecab-ko-dic-2.1.1-20180720.tar.gz
tar zxfv mecab-ko-dic-2.1.1-20180720.tar.gz
cd mecab-ko-dic-2.1.1-20180720
./configure
make
make check
# 아래 실행 후 컴퓨터 비밀번호 입력
sudo make install
```

KoNLP 패키지 설치 및 테스트

KoNLP는 가장 유명한 형태소 분석기입니다.

```
install.packages("KoNLP")  
library(KoNLP)  
  
test <- "한글 테스트 입니다."  
extractNoun(test)
```

아래 결과가 나오면 잘 동작하는 것입니다.

```
[1] "한글" "테스트"
```

KoSpacing 패키지 설치 및 테스트

KoSpacing 패키지는 자동 띄어쓰기 기능을 제공합니다.

```
install.packages("remotes")  
library(remotes)  
install_github('mrchypark/KoSpacing')  
library(KoSpacing)  
set_env()  
spacing("김형호영화시장분석가는'1987'의네이버영화정보네티즌10점평에서언급된단어들을지난해12월27일부터올해1월10일")
```

아래 결과가 나오면 성공입니다.

```
[1] "김형호 영화시장 분석가는 '1987'의 네이버 영화 정보 네티즌 10점 평에서 언급된 단어들을 지난해 12월 27일부
```

RcppMecab 패키지 설치 및 테스트

RcppMecab은 빠른 형태소 분석기입니다.

```
library(remotes)
install_github("junhewk/RcppMeCab")
library(RcppMeCab)

test <- "한글 테스트 입니다."
pos(test)
```

아래 결과가 나와야 합니다.

```
$`한글 테스트 입니다.`
[1] "한글/NNG"      "테스트/NNG"    "입니다/VCP+EF" ". /SF"
```

추가 패키지 설치

tidytext 패키지는 텍스트를 tidy하게 다룰 수 있게 해줍니다.

presidentSpeechKr 패키지는 역대 대통령 연설문 텍스트를 사용할 수 있게 해주는 패키지입니다.

```
install.packages("tidytext")  
install_github("forkonlp/presidentSpeechKr")  
  
library(tidytext)  
library(presidentSpeechKr)
```

위 패키지들은 패키지를 불러오는 것이 동작하는 것을 확인해주세요.

과제

1. 앞의 가이드에 따라 총 5개 패키지를 설치해 주세요.
2. 아래 코드를 R 콘솔에서 실행하여 스크린샷을 제출해 주세요.

```
library(KoNLP)
library(tidytext)
library(presidentSpeechKr)
library(KoSpacing)
library(RcppMeCab)

test <- "한글테스트입니다."
extractNoun(test)
spacing(test)
pos(test)

test <- spacing(test)
extractNoun(test)
```