

# Package ‘KoNLP’

March 20, 2012

**Maintainer** Heewon Jeon <madjakarta@gmail.com>

**License** GPL-3

**Title** Korean NLP Package

**Author** Heewon Jeon

**Description** Korean language processing package. Morphological analyzer, POS tagger, Keystroke converter, Hangul automata...

**SystemRequirements** Java (>= 1.5)

**URL** <https://github.com/haven-jeon/KoNLP>

**Version** 0.7

**Repository** CRAN

**Date/Publication** 2012-03-20 06:54:24

**Date** 2011-10-12

**Encoding** UTF-8

**Depends** R (>= 2.14.0), rJava (>= 0.9-0), bitops (>= 1.0), utils (>= 2.14.0)

**Collate** 'KoAnalyzerRun.R' 'onLoad.R' 'manageDic.R'

## R topics documented:

backupUsrDic . . . . .	2
convertHangulStringToJamos . . . . .	3
convertHangulStringToKeyStrokes . . . . .	3
convertTag . . . . .	4
detectInputEncoding . . . . .	4
doKoMorph . . . . .	5
extractNoun . . . . .	5
HangulAutomata . . . . .	6
is.hangul . . . . .	6

is.jamo . . . . .	7
mergeUserDic . . . . .	7
MorphAnalyzer . . . . .	8
reloadAllDic . . . . .	8
restoreUsrDic . . . . .	9
SimplePos09 . . . . .	10
SimplePos22 . . . . .	10

<b>Index</b>	<b>11</b>
--------------	-----------

---

backupUsrDic	<i>use for backup current dic_user.txt</i>
--------------	--

---

## Description

Utility function for backup dic\_user.txt file to backup directory.

## Usage

```
backupUsrDic(ask = TRUE)
```

## Arguments

ask	ask to confirm backup
-----	-----------------------

## Examples

```
## Not run:
## This codes can not be run if you don't have encoding system which can en/decode Hangul(ex) CP949, EUC-KR, UTF-8).
dicpath <- paste(system.file(package="KoNLP"), "/dics/data/kE/dic_user2.txt", sep="")
newdic <- read.table(dicpath, sep="\t", header=FALSE, fileEncoding="UTF-8", stringsAsFactors=FALSE)
mergeUserDic(newdic)
## backup merged new dictionary
backupUsrDic(ask=FALSE)
## restore from backup directory
restoreUsrDic(ask=FALSE)
## reloading new dictionary
reloadAllDic()
## End(Not run)
```

---

convertHangulStringToJamos  
*conversion function Hangul string to Jamos*

---

### Description

convert Hangul sentence to Jamos. Example will be shown in [github wiki](#).

### Usage

```
convertHangulStringToJamos(hangul)
```

### Arguments

hangul	hangul string
--------	---------------

### Value

Jamo sequences

---

convertHangulStringToKeyStrokes  
*conversion function Hangul string to keyStrokes*

---

### Description

Function can convert Hangul string to Keystrokes. Example will be shown in [github wiki](#).

### Usage

```
convertHangulStringToKeyStrokes(hangul,  
isFullwidth = TRUE)
```

### Arguments

hangul	hangul sentence
isFullwidth	specify returned character will be Fullwidth ASCII or Halfwidth ASCII

### Value

Keystroke sequence

---

convertTag                    *tag name converter*

---

**Description**

only support tag conversion between KAIST and Sejong tag set.

**Usage**

```
convertTag(fromTag, toTag, tag)
```

**Arguments**

fromTag	tag set name to convert from
toTag	desired tag set name
tag	tag name to search

---

detectInputEncoding    *Rough encoding detection function*

---

**Description**

function to be used for file or raw vector encoding detection. This is for internal use.

**Usage**

```
detectInputEncoding(charinput)
```

**Arguments**

charinput	charvector
-----------	------------

**Value**

encoding names of rawinput.

---

doKoMorph	<i>morphological analysis function for Hangul</i>
-----------	---

---

**Description**

morphological analyze the sentence uses lucene korean analyzer. Example will be shown in [github wiki](#).

**Usage**

```
doKoMorph(sentence)
```

**Arguments**

sentence	input
----------	-------

**Value**

stem of sentence

---

extractNoun	<i>Noun extractor for Hangul</i>
-------------	----------------------------------

---

**Description**

extract Nouns from Korean sentence uses Hannanum analyzer. see detail in [Hannanum](#). Example will be shown in [github wiki](#).

**Usage**

```
extractNoun(sentence)
```

**Arguments**

sentence	input
----------	-------

**Value**

Noun of sentence

---

HangulAutomata	<i>do Hangul automata</i>
----------------	---------------------------

---

**Description**

function to be used for converting to complete Hangul syllables from Jamo or Keystrokes. Example will be shown in [github wiki](#).

**Usage**

```
HangulAutomata(input, isKeystroke = F, isForceConv = F)
```

**Arguments**

input	to be processed mostly Jamo sequences
isKeystroke	boolean parameter to check input is keystroke or Jamo sequences
isForceConv	boolean parameter to force converting if input is not valid Jamo or keystroke sequences.

**Value**

complete Hangul syllable

---

is.hangul	<i>check if sentence is all Hangul</i>
-----------	--

---

**Description**

Function checks if each character is Hangul or Jamo. Example will be shown in [github wiki](#).

**Usage**

```
is.hangul(sentenceU8)
```

**Arguments**

sentenceU8	input characters(must be UTF-8)
------------	---------------------------------

**Value**

TRUE or FALSE

---

is.jamo	<i>check if sentence is all Jamo</i>
---------	--------------------------------------

---

**Description**

Function checks with each character is Jamo. Example will be shown in [github wiki](#).

**Usage**

```
is.jamo(sentenceU8)
```

**Arguments**

sentenceU8      input characters(must be UTF-8)

**Value**

TRUE or FALSE

---

mergeUserDic	<i>"dic_user.txt" merging function</i>
--------------	--

---

**Description**

merging current dic\_user.txt with new dictionary.

**Usage**

```
mergeUserDic(newUserDic, append = TRUE, verbose = FALSE)
```

**Arguments**

newUserDic      new user dictionary as data.frame  
 append          append or replacing  
 verbose         see detail error logs

**Examples**

```
## Not run:
## This codes can not be run if you don't have encoding system which can en/decode Hangul(ex) CP949, EUC-KR, UTF-8).
dicpath <- paste(system.file(package="KoNLP"), "/dics/data/kE/dic_user2.txt", sep="")
newdic <- read.table(dicpath, sep="\t", header=FALSE, fileEncoding="UTF-8", stringsAsFactors=FALSE)
mergeUserDic(newdic)
## backup merged new dictionary
backupUsrDic(ask=FALSE)
## restore from backup directory
```

```
restoreUsrDic(ask=FALSE)
## reloading new dictionary
reloadAllDic()
## End(Not run)
```

---

MorphAnalyzer

*Hannanum morphological analyzer interface function*

---

### Description

Do the morphological analysis, not doing pos tagging uses Hannanum analyzer. see details in [Hannanum](#). Example will be shown in [github wiki](#).

### Usage

```
MorphAnalyzer(sentence)
```

### Arguments

sentence      input

### Value

result of analysis

---

reloadAllDic

*reload all Hannanum analyzer dictionary*

---

### Description

Mainly, user dictionary reloading for Hannanum Analyzer. If you want to update user dictionary on KoNLP\_pkg\_dir/inst/dics/data/kE/dic\_user.txt, need to execute this function after editing dic.

### Usage

```
reloadAllDic()
```

**Examples**

```
## Not run:
## This codes can not be run if you don't have encoding system which can en/decode Hangul(ex) CP949, EUC-KR, UTF-8)
dicpath <- paste(system.file(package="KoNLP"), "/dics/data/kE/dic_user2.txt", sep="")
newdic <- read.table(dicpath, sep="\t", header=FALSE, fileEncoding="UTF-8", stringsAsFactors=FALSE)
mergeUserDic(newdic)
## backup merged new dictionary
backupUsrDic(ask=FALSE)
## restore from backup directory
restoreUsrDic(ask=FALSE)
## reloading new dictionary
reloadAllDic()
## End(Not run)
```

---

restoreUsrDic	<i>use for restoring backedup dic_user.txt</i>
---------------	--

---

**Description**

Utility function for restoring dic\_user.txt file to dictionary directory.

**Usage**

```
restoreUsrDic(ask = TRUE)
```

**Arguments**

ask                    ask to confirm backup

**Examples**

```
## Not run:
## This codes can not be run if you don't have encoding system which can en/decode Hangul(ex) CP949, EUC-KR, UTF-8).
dicpath <- paste(system.file(package="KoNLP"), "/dics/data/kE/dic_user2.txt", sep="")
newdic <- read.table(dicpath, sep="\t", header=FALSE, fileEncoding="UTF-8", stringsAsFactors=FALSE)
mergeUserDic(newdic)
## backup merged new dictionary
backupUsrDic(ask=FALSE)
## restore from backup directory
restoreUsrDic(ask=FALSE)
## reloading new dictionary
reloadAllDic()
## End(Not run)
```

---

SimplePos09

*POS tagging by using 9 KAIST tags*

---

**Description**

Do pos tagging using 9 tags uses Hannanum analyzer. see details in [Hannanum](#). Example will be shown in [github wiki](#).

**Usage**

```
SimplePos09(sentence)
```

**Arguments**

sentence          input

**Value**

Noun of sentence

---

SimplePos22

*POS tagging by using 22 KAIST tags*

---

**Description**

Do POS tagging using 22 tags uses Hannanum analyzer. see details in [Hannanum](#). Example will be shown in [github wiki](#).

**Usage**

```
SimplePos22(sentence)
```

**Arguments**

sentence          input

**Value**

result of analysis

# Index

backupUsrDic, [2](#)

convertHangulStringToJamos, [3](#)

convertHangulStringToKeyStrokes, [3](#)

convertTag, [4](#)

detectInputEncoding, [4](#)

doKoMorph, [5](#)

extractNoun, [5](#)

HangulAutomata, [6](#)

is.hangul, [6](#)

is.jamo, [7](#)

mergeUserDic, [7](#)

MorphAnalyzer, [8](#)

reloadAllDic, [8](#)

restoreUsrDic, [9](#)

SimplePos09, [10](#)

SimplePos22, [10](#)