

Package: HanNLP (via r-universe)

October 1, 2026

Type Package

Title Korean Morphological Analysis and Hangul Utilities

Version 0.1.1

Maintainer Jonghyeon Park <jng.hyn.park@gmail.com>

Description Provides Korean morphological analysis, part-of-speech (POS) tagging, noun extraction, Hangul (Korean script) conversion utilities, concordance search, mutual information statistics, and user dictionary management tools for Korean text research. The morphological analyzer backend is derived from the 'KoNLP' package and the 'HanNanum' analyzer, and is reimplemented in native C so that no Java runtime is required. Bundled dictionary and statistical resources allow analysis to work out of the box, and optional dictionaries can be loaded at runtime from the 'Sejong' and 'NIADic' packages when they are installed and loaded.

License GPL-3

Copyright See inst/COPYRIGHTS for bundled third-party code and data.

URL <https://github.com/ShapeLayer/HanNLP>

BugReports <https://github.com/ShapeLayer/HanNLP/issues>

Encoding UTF-8

Depends R (>= 4.0.0)

Imports utils

NeedsCompilation yes

Roxygen list(markdown = TRUE)

Repository <https://shapelayer.r-universe.dev>

Date/Publication 2026-10-01 01:41:22 UTC

RemoteUrl <https://github.com/shapelayer/hannlp>

RemoteRef HEAD

RemoteSha c1ecfa314d1b503a9d6d7df8e754db23ddfdc66c

Contents

backupUsrDic	2
buildDictionary	3
concordance_file	4
concordance_str	4
convertHangulStringToJamos	5
convertHangulStringToKeyStrokes	5
convertTag	6
editweights	6
extractNoun	7
get_dictionary	7
HangulAutomata	8
is.ascii	8
is.hangul	9
is.jaeum	9
is.jamo	10
is.moeum	10
KtoS	11
mergeUserDic	11
MorphAnalyzer	12
mutualinformation	12
reloadAllDic	13
reloadUserDic	13
restoreUsrDic	14
SimplePos09	14
SimplePos22	15
statDic	15
StoK	16
tags	16
useNIADic	17
useSejongDic	17
useSystemDic	18

Index	19
--------------	-----------

backupUsrDic	<i>Backup current user dictionary</i>
--------------	---------------------------------------

Description

Backup the writable HanNLP user dictionary.

Usage

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

Arguments

ask ask to confirm backup

Details

Before writing dictionary files, explicitly set the environment variable HANNLP_USER_DATA_DIR to the desired directory. There is no default writable directory. Backups are stored in its ‘backup’ subdirectory. For temporary use, choose a directory under tempdir().

Value

invisible logical

buildDictionary	<i>Build HanNLP user dictionary</i>
-----------------	-------------------------------------

Description

Build the writable HanNLP user dictionary from user-supplied entries and optional external sources. The Sejong dictionary is read from the installed Sejong package or the path specified by HANNLP_SEJONG_HANDIC. The woorimalsam and insighter dictionaries require the optional NIADic package to be installed and loaded (library(NIADic)) in the current session; HanNLP does not bundle NIADic data.

Usage

```
buildDictionary(  
  ext_dic = "",  
  category_dic_nms = "",  
  user_dic = data.frame(),  
  replace_usr_dic = FALSE,  
  verbose = FALSE  
)
```

Arguments

ext_dic optional external dictionary names: sejong, woorimalsam, or insighter
category_dic_nms retained for KoNLP API compatibility
user_dic user dictionary data.frame with term and KAIST tag columns
replace_usr_dic replace existing user dictionary instead of appending
verbose print detail progress

Details

Before writing dictionary files, explicitly set the environment variable HANNLP_USER_DATA_DIR to the desired directory. There is no default writable directory. Backups are stored in its ‘backup’ subdirectory. For temporary use, choose a directory under tempdir().

Value

invisible data.frame of the resulting user dictionary

concordance_file	<i>Concordance for input text file</i>
------------------	--

Description

Return concordance text for an input file.

Usage

```
concordance_file(filename, pattern, encoding = getOption("encoding"), span = 5)
```

Arguments

filename	file name
pattern	pattern of central words
encoding	file encoding
span	number of characters to include around each match

Value

character vector of matched concordance strings

concordance_str	<i>Concordance for input text vector</i>
-----------------	--

Description

Return concordance text for an input pattern and span.

Usage

```
concordance_str(string, pattern, span = 5)
```

Arguments

string	input text as character vector or single character
pattern	pattern of central words
span	number of characters to include around each match

Value

list of matched concordance strings, with empty elements removed

```
convertHangulStringToJamos
```

Convert Hangul string to Jamos

Description

Convert Hangul syllables to compatibility Jamo sequences.

Usage

```
convertHangulStringToJamos(hangul)
```

Arguments

hangul	Hangul string
--------	---------------

Value

Jamo sequences

Examples

```
hangul <- intToUtf8(c(0xd55c, 0xae00))  
convertHangulStringToJamos(hangul)
```

```
convertHangulStringToKeyStrokes
```

Convert Hangul string to keystrokes

Description

Convert Hangul syllables to two-beolsik keystroke sequences.

Usage

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

Arguments

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

Value

Keystroke sequence

convertTag	<i>Tag name converter</i>
------------	---------------------------

Description

Convert tags between KAIST and Sejong tag sets.

Usage

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

Arguments

fromTag	tag set name to convert from, either K or S
toTag	desired tag set name, either K or S
tag	tag name to search

Value

converted tag vector

Examples

```
convertTag("K", "S", "ncn")
```

editweights	<i>Keystroke misspell cost table</i>
-------------	--------------------------------------

Description

KoNLP-compatible edit weight metadata. HanNLP does not use this table in runtime algorithms yet.

Format

A 66 by 66 numeric matrix of edit (substitution) costs indexed by keystroke characters on both dimensions.

Value

A 66 by 66 numeric matrix of keystroke edit costs.

Source

Derived from the edit weight table distributed with 'KoNLP'.

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

Description

Extract nouns from Korean sentences using the native HanNLP analyzer backend.

Usage

```
extractNoun(sentences, autoSpacing = FALSE)
```

Arguments

sentences	input character vector
autoSpacing	retained for KoNLP API compatibility; currently ignored

Value

character vector for one sentence, list for multiple sentences

Examples

```
local({
  old <- Sys.getenv("HANNLP_USER_DATA_DIR", unset = NA)
  tmp <- tempfile("hannlp-user-data-")
  on.exit({
    if (is.na(old)) Sys.unsetenv("HANNLP_USER_DATA_DIR") else Sys.setenv(HANNLP_USER_DATA_DIR = old)
    unlink(tmp, recursive = TRUE, force = TRUE)
  })
  Sys.setenv(HANNLP_USER_DATA_DIR = tmp)
  text <- intToUtf8(c(0xd55c, 0xae00, 0x20, 0xd615, 0xd0dc, 0xc18c))
  extractNoun(text)
})
```

get_dictionary	<i>Get Dictionary</i>
----------------	-----------------------

Description

Return a HanNLP dictionary as a data.frame. The external Sejong dictionary requires an installed source package or a configured file path. The NIADic-based dictionaries require the loaded optional NIADic package.

Usage

```
get_dictionary(dic_name)
```

Arguments

dic_name one of user_dic, system_dic, analyzed_dic, sejong, woorimalsam, or insighter. The woorimalsam and insighter dictionaries require the optional NIADic package to be installed and loaded (library(NIADic)) in the current session; HanNLP does not bundle NIADic data.

Value

data.frame with dictionary terms and tags

HangulAutomata	<i>Compose Hangul syllables from Jamo or keystrokes</i>
----------------	---

Description

Compose complete Hangul syllables from compatibility Jamo or two-beolsik keystroke sequences.

Usage

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

Arguments

input input Jamo sequences or two-beolsik keystrokes
 isKeystroke whether input is two-beolsik keystrokes instead of Jamo
 isForceConv force conversion for incomplete sequences

Value

complete Hangul syllable string

is.ascii	<i>Check if sentence is all ASCII</i>
----------	---------------------------------------

Description

Check whether each input string contains only ASCII characters.

Usage

```
is.ascii(sentence)
```

Arguments

sentence input characters

Value

logical vector

is.hangul

Check if sentence is all Hangul

Description

Check whether each input string contains only Hangul syllables or compatibility Jamo.

Usage

```
is.hangul(sentence)
```

Arguments

sentence input characters

Value

logical vector

Examples

```
hangul <- intToUtf8(c(0xd55c, 0xae00))  
is.hangul(c(hangul, "abc"))
```

is.jaeum

Check if sentence is all Jaeum

Description

Check whether each input string contains only Korean consonant Jamo.

Usage

```
is.jaeum(sentence)
```

Arguments

sentence input characters

Value

logical vector

`is.jamo`*Check if sentence is all Jamo*

Description

Check whether each input string contains only compatibility Jamo.

Usage

```
is.jamo(sentence)
```

Arguments

sentence input characters

Value

logical vector

`is.moeum`*Check if sentence is all Moeum*

Description

Check whether each input string contains only Korean vowel Jamo.

Usage

```
is.moeum(sentence)
```

Arguments

sentence input characters

Value

logical vector

KtoS	<i>KAIST tag to Sejong tag</i>
------	--------------------------------

Description

KoNLP-compatible mapping from KAIST tags to Sejong tags.

Format

A named character vector whose names are KAIST tag names and whose values are the corresponding Sejong tag names.

Value

A named character vector mapping KAIST tags to Sejong tags.

Source

Derived from the tag mapping distributed with 'KoNLP'.

mergeUserDic	<i>Append or replace the HanNLP user dictionary</i>
--------------	---

Description

Append to or replace the writable HanNLP user dictionary.

Usage

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

Arguments

<code>newUserDic</code>	new user dictionary as data.frame
<code>append</code>	append to existing dictionary or replace it
<code>verbose</code>	retained for KoNLP API compatibility
<code>ask</code>	ask to backup current dictionary

Details

Before writing dictionary files, explicitly set the environment variable HANNLP_USER_DATA_DIR to the desired directory. There is no default writable directory. Backups are stored in its 'backup' subdirectory. For temporary use, choose a directory under `tempdir()`.

Value

invisible data.frame of the resulting user dictionary

MorphAnalyzer	<i>Hannanum morphological analyzer interface function</i>
---------------	---

Description

Analyze Korean sentences using the native HanNLP analyzer backend.

Usage

```
MorphAnalyzer(sentences, autoSpacing = FALSE)
```

Arguments

sentences	input character vector
autoSpacing	retained for KoNLP API compatibility; currently ignored

Value

named list of eojeol to morpheme/tag strings for one sentence, list for multiple sentences

Examples

```
text <- intToUtf8(c(0xd55c, 0xae00, 0x20, 0xd615, 0xd0dc, 0xc18c))
MorphAnalyzer(text)
```

mutualinformation	<i>Mutual information for input text</i>
-------------------	--

Description

Return mutual information or t-scores for adjacent word bigrams.

Usage

```
mutualinformation(text, query = "", method = c("mutual", "tscores"))
```

Arguments

text	input character vector
query	term to filter bigrams
method	calculation method, either mutual or tscores

Value

named numeric vector

reloadAllDic	<i>Reload all dictionaries</i>
--------------	--------------------------------

Description

HanNLP loads dictionaries per analysis call, so this function is retained for KoNLP API compatibility.

Usage

```
reloadAllDic()
```

Value

invisible TRUE

reloadUserDic	<i>Reload user dictionaries</i>
---------------	---------------------------------

Description

Compatibility wrapper for reloading user dictionaries.

Usage

```
reloadUserDic(whichDics)
```

Arguments

whichDics	character vector of dictionary consumers
-----------	--

Value

invisible TRUE

restoreUsrDic	<i>Restore backed up user dictionary</i>
---------------	--

Description

Restore the writable HanNLP user dictionary from backup.

Usage

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

Arguments

ask	ask to confirm restore
-----	------------------------

Details

Before writing dictionary files, explicitly set the environment variable HANNLP_USER_DATA_DIR to the desired directory. There is no default writable directory. Backups are stored in its ‘backup’ subdirectory. For temporary use, choose a directory under tempdir().

Value

invisible logical

SimplePos09	<i>POS tagging by using 9 KAIST tags</i>
-------------	--

Description

Tag Korean sentences using 9 KAIST tags with the native HanNLP analyzer backend.

Usage

```
SimplePos09(sentences, autoSpacing = FALSE)
```

Arguments

sentences	input character vector
autoSpacing	retained for KoNLP API compatibility; currently ignored

Value

named list of eojeol to morpheme/tag strings for one sentence, list for multiple sentences

SimplePos22

*POS tagging by using 22 KAIST tags***Description**

Tag Korean sentences using 22 KAIST tags with the native HanNLP analyzer backend.

Usage

```
SimplePos22(sentences, autoSpacing = FALSE)
```

Arguments

sentences	input character vector
autoSpacing	retained for KoNLP API compatibility; currently ignored

Value

named list of eojeol to morpheme/tag strings for one sentence, list for multiple sentences

statDic

*Summary of dictionaries***Description**

Show summary, head, and tail of current or backup user dictionaries.

Usage

```
statDic(which = "current", n = 6)
```

Arguments

which	one of current or backup
n	number of rows for head and tail

Value

list containing summary, head, and tail

StoK	<i>Sejong tag to KAIST tag</i>
------	--------------------------------

Description

KoNLP-compatible mapping from Sejong tags to KAIST tags.

Format

A named character vector whose names are Sejong tag names and whose values are the corresponding KAIST tag names.

Value

A named character vector mapping Sejong tags to KAIST tags.

Source

Derived from the tag mapping distributed with 'KoNLP'.

tags	<i>Tag names</i>
------	------------------

Description

KoNLP-compatible KAIST/HanNanum tag names used by HanNLP dictionary utilities.

Format

A named character vector mapping each supported tag name to itself, used to validate tag entries.

Value

A named character vector of supported tag names.

Source

Derived from the KAIST/HanNanum tag set distributed with 'KoNLP'.

useNIADic	<i>Use NIA dictionary</i>
-----------	---------------------------

Description

Replace the writable user dictionary with NIADic entries. HanNLP does not bundle or distribute NIADic data because it is released under CC BY-SA, which is incompatible with the GPL. This function works only when the optional NIADic package is installed and its namespace is already loaded in the current R session (e.g. after `library(NIADic)`); otherwise it errors.

Usage

```
useNIADic(
  which_dic = c("woorimalsam", "insighter"),
  category_dic_nms = "all",
  backup = TRUE
)
```

Arguments

<code>which_dic</code>	NIADic dictionaries to load: woorimalsam, insighter, or both
<code>category_dic_nms</code>	retained for KoNLP API compatibility
<code>backup</code>	backup current dictionary before switching

Details

Before writing dictionary files, explicitly set the environment variable `HANNLP_USER_DATA_DIR` to the desired directory. There is no default writable directory. Backups are stored in its ‘backup’ subdirectory. For temporary use, choose a directory under `tempdir()`.

Value

invisible data.frame of the resulting user dictionary

useSejongDic	<i>Use bundled Sejong-style dictionary</i>
--------------	--

Description

Replace the writable user dictionary with Sejong dictionary entries from an installed Sejong package or `HANNLP_SEJONG_HANDIC`.

Usage

```
useSejongDic(backup = TRUE)
```

Arguments

backup backup current dictionary before switching

Details

Before writing dictionary files, explicitly set the environment variable HANNLP_USER_DATA_DIR to the desired directory. There is no default writable directory. Backups are stored in its ‘backup’ subdirectory. For temporary use, choose a directory under tempdir().

Value

invisible result

useSystemDic	<i>Use system default dictionary</i>
--------------	--------------------------------------

Description

Restore the writable user dictionary to HanNLP’s bundled default user dictionary.

Usage

useSystemDic(backup = TRUE)

Arguments

backup backup current dictionary before switching

Details

Before writing dictionary files, explicitly set the environment variable HANNLP_USER_DATA_DIR to the desired directory. There is no default writable directory. Backups are stored in its ‘backup’ subdirectory. For temporary use, choose a directory under tempdir().

Value

invisible TRUE

Index

- * **datasets**
 - editweights, [6](#)
 - KtoS, [11](#)
 - StoK, [16](#)
 - tags, [16](#)
- backupUsrDic, [2](#)
- buildDictionary, [3](#)
- concordance_file, [4](#)
- concordance_str, [4](#)
- convertHangulStringToJamos, [5](#)
- convertHangulStringToKeyStrokes, [5](#)
- convertTag, [6](#)
- editweights, [6](#)
- extractNoun, [7](#)
- get_dictionary, [7](#)
- HangulAutomata, [8](#)
- is.ascii, [8](#)
- is.hangul, [9](#)
- is.jaeum, [9](#)
- is.jamo, [10](#)
- is.moeum, [10](#)
- KtoS, [11](#)
- mergeUserDic, [11](#)
- MorphAnalyzer, [12](#)
- mutualinformation, [12](#)
- reloadAllDic, [13](#)
- reloadUserDic, [13](#)
- restoreUsrDic, [14](#)
- SimplePos09, [14](#)
- SimplePos22, [15](#)
- statDic, [15](#)
- StoK, [16](#)
- tags, [16](#)
- useNIADic, [17](#)
- useSejongDic, [17](#)
- useSystemDic, [18](#)