Package 'sudachir'

May 9, 2026

Type Package

Title R Interface to 'Sudachi'

Version 0.1.0

Maintainer Shinya Uryu <suika1127@gmail.com>

Description Interface to 'Sudachi' <https://github.com/WorksApplications/Sudachi>, a Japanese morphological analyzer. This is a port of what is available in Python.

License Apache License (>= 2.0)

Encoding UTF-8

SystemRequirements Python (>= 2.7.0)

URL https://github.com/uribo/sudachir

BugReports https://github.com/uribo/sudachir/issues

Imports cli (>= 2.1.0), dplyr (>= 1.0.2), glue (>= 1.4.2), magrittr (>= 1.5), purrr (>= 0.3.4), rlang (>= 0.4.8), reticulate (>= 1.17), tibble (>= 3.0.4), tidyselect (>= 1.1.0)

LazyData true

RoxygenNote 7.1.1

Suggests rstudioapi, testthat

NeedsCompilation no

Author Shinya Uryu [aut, cre] (ORCID: <https://orcid.org/0000-0002-0493-6186>), Akiru Kato [aut]

Repository CRAN

Date/Publication 2020-11-10 15:20:02 UTC

Contents

`create_sudachipy_env` *Create conda env used by sudachir*

Description

Create conda env used by sudachir

Usage

```
create_sudachipy_env(python_version = 3.9)
```

Arguments

`python_version`	Python version to use within conda environment created for installing the SudachiPy It requires version 3.5 or higher.

`form` *Parse tokenized input text*

Description

Parse tokenized input text

Usage

```
form(x, mode, type, pos = TRUE)
```

Arguments

`x`	Input text vectors
`mode`	Select split mode (A, B, C)
`type`	return form. One of the following "surface", "dictionary", "normalized", "reading" or "part_of_speech".
`pos`	Include part of speech information with object name.

Examples

```
## Not run:
form("Tokyo", mode = "B", type = "normalized")
form("Osaka", mode = "B", type = "surface")
form("Hokkaido", mode = "C", type = "part_of_speech")

## End(Not run)
```

install_sudachipy *Install SudachiPy*

Description

Install SudachiPy to Conda virtual environment. As a one-time setup step, you most run `install_sudachipy()` to install all dependencies.

Usage

```
install_sudachipy()
```

Details

`install_sudachipy()` requires Python and Conda to be installed. See https://www.python.org/getit/ and https://docs.conda.io/projects/conda/en/latest/user-guide/install/.

Examples

```
## Not run:
install_sudachipy()

## End(Not run)
```

rebuild_tokenizer *Rebuild tokenizer*

Description

Rebuild tokenizer

Usage

```
rebuild_tokenizer(config_path = NULL)
```

Arguments

`config_path`	Absolute path to `sudachi.json`

Value

Returns a binding to the instance of `<sudachipy.tokenizer.Tokenizer>`.

Examples

```
## Not run:
instance <- rebuild_tokenizer()
tokenizer("Tokyo, Japan", mode = "A", instance)

## End(Not run)
```

`remove_sudachipy` *Remove SudachiPy*

Description

Uninstalls SudachiPy by removing the Conda environment.

Usage

```
remove_sudachipy()
```

Examples

```
## Not run:
install_sudachipy()
remove_sudachipy()

## End(Not run)
```

`tokenizer` *Sudachi tokenizer*

Description

Sudachi tokenizer

Usage

```
tokenizer(x, mode, instance = NULL)
```

Arguments

`x`	Input text vectors
`mode`	Select split mode (A, B, C)
`instance`	This is optional if you already have an instance of `<sudachipy.tokenizer.Tokenizer>` Giving them a predefined instance will speed up their execution.

Examples

```
## Not run:
tokenizer("Tokyo, Japan", mode = "A")

## End(Not run)
```

---

tokenize_to_df *Create tokenizing data.frame using Sudachi*

Description

Create tokenizing data.frame using Sudachi

Usage

```
tokenize_to_df(x, mode, instance = NULL)
```

Arguments

x	Input text vectors
mode	Select split mode (A, B, C)
instance	This is optional if you already have an instance of <sudachipy.tokenizer.Tokenizer> Giving them a predefined instance will speed up their execution.

Examples

```
## Not run:
tokenizer("Tokyo, Japan", mode = "A")

## End(Not run)
```

Index