Package ‘N2H4’
February 25, 2024

Type    Package
Title   Handling Methods for Naver News Text Crawling
Version 0.8.4
Description Provides some functions to get Korean text sample from news articles in Naver which is popular news portal service <https://news.naver.com/> in Korea.
License MIT + file LICENSE
URL https://forkonlp.github.io/N2H4/, https://github.com/forkonlp/N2H4
BugReports https://github.com/forkonlp/N2H4/issues
RoxygenNote 7.2.3
Depends R (>= 3.5.0)
Encoding UTF-8
Suggests testthat, devtools, usethis
Imports rvest, jsonlite, tibble, lifecycle, httr2, magrittr
NeedsCompilation no
Author Chanyub Park [aut, cre] (<https://orcid.org/0000-0001-6474-2570>)
Maintainer Chanyub Park <mrchypark@gmail.com>
Repository CRAN
Date/Publication 2024-02-25 16:20:02 UTC

R topics documented:

getAllComment ................................................. 2
getAllCommentHistory ....................................... 2
category ....................................................... 3
category ....................................................... 3
gCategory ................................................... 3
gCommentHistory ............................................ 4
gCommentHistory ............................................ 4
gContent ..................................................... 5
getMainCategory ............................................. 5
getSubCategory .............................................. 6

Index 7
getAllComment 

**Description**

Get all comments from the provided news article url on naver

**Usage**

ggetAllComment(turl)

**Arguments**

turl  character. News article on `Naver` such as <https://n.news.naver.com/mnews/article/023/0003712918>. News article url that is not on Naver.com domain will generate an error.

**Details**

Works just like getComment, but this function executed in a fashion where it finds and extracts all comments from the given url.

**Value**

a [tibble][tibble::tibble-package]

**Examples**

```r
## Not run:
ggetAllComment("https://n.news.naver.com/mnews/article/214/0001195110")
## End(Not run)
```

ggetAllCommentHistory

**Description**

Get All Comment History

**Usage**

ggetAllCommentHistory(turl, commentNo)

**Arguments**

turl  character. News article on `Naver` such as <https://n.news.naver.com/mnews/article/001/0009205077?sid=102>. News article url that is not on Naver.com domain will generate an error.

commentNo  Parent Comment No.
**getCategory**

### Value

a [tibble][tibble::tibble-package]

### Examples

```r
## Not run:
getAllComment("https://n.news.naver.com/mnews/article/214/0001195110?sid=103")

## End(Not run)
```

---

**getDescription**

<table>
<thead>
<tr>
<th>News Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
</tr>
<tr>
<td>News Category</td>
</tr>
</tbody>
</table>

### Usage

`getCategory(fresh = FALSE)`

### Arguments

- `fresh`: get data from online. Default is FALSE using cached built-in data.

---

**getComment**

### Description

Get naver news comments. if you want to get data only comment, enter command like below.

getComment(url)$result$commentList[[1]]

### Usage

`getComment(turl, count = 10, type = c("df", "list"))`

### Arguments

- `count`: is a number of comments. Default is 10. "all" works to get all comments.
- `type`: type return df or list. Default is df. df return part of data not all.

### Value

a [tibble][tibble::tibble-package]
getCommentHistory

Description
Get naver news comments on user histories.

Usage
getCommentHistory(turl, commentNo, count = 10, type = c("df", "list"))

Arguments
- turl: character. News article on 'Naver' such as <https://n.news.naver.com/mnews/article/001/0009205077?sid=102>. News article url that is not on Naver.com domain will generate an error.
- commentNo: Parent Comment No.
- count: is a number of comments. Default is 10. "all" works to get all comments.
- type: type return df or list. Default is df. df return part of data not all.

Value
a [tibble][tibble::tibble-package]

Examples
## Not run:
cno <- getComment("https://n.news.naver.com/mnews/article/421/0002484966?sid=100")
getCommentHistory("https://n.news.naver.com/mnews/article/421/0002484966?sid=100", cno$commentNo[1])
## End(Not run)
**getContent**

**Get Content**

**Description**

Get naver news content from links.

**Usage**

```r
g getContent(
    turl,
    col = c("url", "original_url", "section", "datetime", "edittime", "press", "title", "body")
)
```

**Arguments**

- `turl` is naver news link.
- `col` is what you want to get from news. Default is all.

**Value**

a `[tibble]`

**Examples**

```r
# Not run:
g getContent("https://n.news.naver.com/mnews/article/214/0001195110?sid=103")
```

**getMainCategory**

**Get News Main Categories**

**Description**

Get naver news main category names and ids recently.

**Usage**

```r
g getMainCategory()
```

**Value**

a `[tibble]`
getSubCategory

Examples

## Not run:
getMainCategory()

## End(Not run)

getSubCategory | Get News Sub Categories

Description

Get naver news sub category names and urls recently.

Usage

getSubCategory(sid1 = 100)

Arguments

sid1            Main category id in naver news url. Only 1 value is possible. Default is 100
means Politics.

Value

a [tibble][tibble::tibble-package]

Examples

## Not run:
getSubCategory(100)
getSubCategory(100, FALSE)

## End(Not run)
Index

getAllComment, 2
getAllCommentHistory, 2
getCategory, 3
getComment, 3
getCommentHistory, 4
g getContent, 5
getMainCategory, 5
getSubCategory, 6