Package ‘N2H4’

October 12, 2022

Type  Package
Title  Handling Methods for Naver News Text Crawling
Version  0.8.2
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://github.com/forkonlp/N2H4
BugReports  https://github.com/forkonlp/N2H4/issues
RoxygenNote  7.2.1
Depends  R (>= 3.5.0)
Encoding  UTF-8
Suggests  testthat, devtools, usethis
Imports  rvest, jsonlite, tibble, purrr, 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  2022-09-15 09:20:02 UTC

R topics documented:

getAllComment .................................................. 2
getAllCommentHistory ....................................... 2
getCategory .................................................... 3
getComment ..................................................... 3
gCommentHistory ............................................... 4
getContent ....................................................... 5
getMainCategory ............................................... 5
getMaxPageNum ................................................. 6
getSubCategory ............................................... 6
getUrlList ....................................................... 7
getAllComment

**Description**

Get all comments from the provided news article url on naver

**Usage**

```r
getAllComment(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:
getAllComment("https://n.news.naver.com/mnews/article/214/0001195110")
```

## End(Not run)

getAllCommentHistory

**Description**

Get All Comment History

**Usage**

```r
ggetAllCommentHistory(turl, commentNo)
```

**Examples**

```r
## Not run:
ggetAllCommentHistory("https://n.news.naver.com/mnews/article/214/0001195110", 1)
```
**getCategory**

**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.

**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**

**Usage**

```r
generateCategory(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.

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

**Usage**

```r
generateComment(turl, count = 10, type = c("df", "list"))
```
getCommentHistory

Arguments

turl character. News article on 'Naver' such as <https://n.news.naver.com/mnews/article/001/0009205077?sid=102>. News articl 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. Defualt 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")

cno$commentNo[1]

## End(Not run)

---

getCommentHistory Get Comment History

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 articl 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. Defualt is df. df return part of data not all.

Value

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

Examples

## Not run:

getComment("https://n.news.naver.com/mnews/article/421/0002484966?sid=100")

cno$commentNo[1]

## End(Not run)
**getContent**  

---

**getDescription**  

Get Naver news content from links.

**Usage**

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

**Arguments**

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

**Value**

A `tibble` [tibble::tibble-package](https://CRAN.R-project.org/package=tibble)

**Examples**

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

---

**getMainCategory**  

---

**getDescription**  

Get Naver news main category names and ids recently.

**Usage**

```r
getMainCategory()
```

**Value**

A `tibble` [tibble::tibble-package](https://CRAN.R-project.org/package=tibble)
getMaxPageNum  
Get Max Page Number

Description
Get Max Page Number

Usage
getMaxPageNum(turl, max = 100)

Arguments

turl
is target url include sid1, sid2, date like below. <http://news.naver.com/main/list.nhn?sid2=265&sid1=100&mid=shm&mode=LS2D&date=20161102>

max
is also interval to try max page number is numeric. Default is 100.

Value
Get numeric

Examples

## Not run:
getMaxPageNum("https://news.naver.com/main/list.naver?mode=LS2D&mid=shm&sid1=103&sid2=376")

## End(Not run)

getSubCategory  
Get News Sub Categories

Description
Get naver news sub category names and urls recently.

Usage
getSubCategory(sid1 = 100)
getUrlList

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

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

## End(Not run)
```

getUrlList  Get Url List By Category

Description

Get naver news titles and links from target url.

Usage

```r
getUrlList(turl, col = c("titles", "links"))
```

Arguments

**turl** is target url naver news.

**col** is what you want to get from news. Default is all.

Value

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

Examples

```r
## Not run:
getUrlList("https://news.naver.com/main/list.naver?mode=LS2D&mid=shm&sid1=103&sid2=376")

## End(Not run)
```
Index

getAllComment, 2
getAllCommentHistory, 2
getCategory, 3
getComment, 3
getCommentHistory, 4
getContent, 5
getMainCategory, 5
getMaxPageNum, 6
getSubCategory, 6
getUrlList, 7