

Package ‘DNH4’

March 9, 2022

Type Package

Title Crawling for Daum News Text

Version 0.1.12

Date 2022-03-06

Description Provides some utils to get Korean text sample from news articles in Daum which is popular news portal service in Korea.

URL <https://forkonlp.github.io/DNH4/>,
<https://github.com/forkonlp/DNH4/>

BugReports <https://github.com/forkonlp/DNH4/issues>

RoxygenNote 7.1.2

Language r(>=3.3.0)

Encoding UTF-8

Imports httr, rvest, tidyr, tibble

Suggests httpptest, testthat

License MIT + file LICENSE

Depends R (>= 2.10)

NeedsCompilation no

Author Chanyub Park [aut, cre] (<<https://orcid.org/0000-0001-6474-2570>>)

Maintainer Chanyub Park <mrchypark@gmail.com>

Repository CRAN

Date/Publication 2022-03-09 08:20:16 UTC

R topics documented:

getAllComment	2
getComment	2
getContent	3
getMainCategory	3
getMaxPageNum	4
getSubCategory	4
getUrlList	5

Index**6**

getAllComment	<i>Get All Comment</i>
---------------	------------------------

Description

Get daum news comments

Usage

```
getAllComment(turl, sort = c("RECOMMEND", "LATEST"))
```

Arguments

turl	like 'http://v.media.daum.net/v/20161117210603961'.
sort	you can select RECOMMEND, LATEST. RECOMMEND is Default.

Value

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

getComment	<i>Get Comment</i>
------------	--------------------

Description

Get daum news comments

Usage

```
getComment(
  turl,
  limit = 10,
  offset = 0,
  parentId = 0,
  sort = c("RECOMMEND", "LATEST"),
  type = c("df", "list")
)
```

Arguments

turl	like 'http://v.media.daum.net/v/20161117210603961'.
limit	is number of comment. Default is 10.
offset	is comment number of start. Default is 0.
parentId	Default is 0.
sort	you can select RECOMMEND, LATEST. RECOMMEND is Default.
type	return type. Default is tibble. It may sometimes warn message.

Value

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

getContent	<i>Get Content</i>
------------	--------------------

Description

Get daum news content from links.

Usage

```
getContent(turl = url)
```

Arguments

turl is daum news link.

Value

a [tibble][tibble::tibble-package] (url,datetime,press,title,content).

getMainCategory	<i>Get News Main Categories</i>
-----------------	---------------------------------

Description

Get daum news main category names and ids recently.

Usage

```
getMainCategory(fresh = FALSE)
```

Arguments

fresh If TRUE, get data from internet. Default is FALSE which is return with cache.

Value

Get data.frame(chr:cate_name, chr:url).

Examples

```
getMainCategory()
```

getMaxPageNum	<i>Get Max Page Number</i>
---------------	----------------------------

Description

Get Max Page Number

Usage

```
getMaxPageNum(turl = url)
```

Arguments

turl	is target url include breakingnews, category url, date without regDate like below. 'https://news.daum.net/breakingnews/politics/administration?regDate=20220305'
------	---

Value

Get numeric

getSubCategory	<i>Get News Sub Categories</i>
----------------	--------------------------------

Description

Get daum news sub category names and urls recently.

Usage

```
getSubCategory(categoryUrl = "society", fresh = FALSE)
```

Arguments

categoryUrl	Main category url in daum news. Only 1 value is passible. Default is society.
fresh	If TRUE, get data from internet. Default is FALSE which is return with cache.

Value

Get data.frame(chr:sub_cate_name, chr:url).

Examples

```
getSubCategory()  
getSubCategory("politics")
```

getUrlList

Get Url List

Description

Get daum news titles and links from target url.

Usage

```
getUrlList(turl = url)
```

Arguments

turl is target url daum news.

Value

a [tibble][tibble::tibble-package](news_title, news_links).

Index

[getAllComment](#), [2](#)
[getComment](#), [2](#)
[getContent](#), [3](#)
[getMainCategory](#), [3](#)
[getMaxPageNum](#), [4](#)
[getSubCategory](#), [4](#)
[getUrlList](#), [5](#)