Package 'AnalyseMedia'

July 12, 2016

Type Package

Title AnalyseMedia

Version 0.11

Author Martin Badicke

Maintainer <martin.badicke@gmail.com>

Depends R (>= 3.2.4), SPARQL (>= 1.16), dplyr (>= 0.4.3), xml2 (>= 0.1.2), rvest (>= 0.3.1), stringr (>= 1.0.0), tidyr (>= 0.4.1), rmarkdown (>= 0.9.5), knitr (>= 1.12.3)

Imports

ByteCompile TRUE

Description AnalyseMedia

License closed

RoxygenNote 5.0.1

NeedsCompilation no

R topics documented:

aggStatTab *Build aggregation table*

Description

Selection of aggrageted column informations.

Usage

```
aggStatTab(.dat, selectSearchTerm, type = "aggregatedArticleStatistic")
```

Arguments

.dat	The object with article data, preferably the output from function cleanData.
selectSearchTerm	The search-term for which to create the table (defining the relevent rows of dat).
type	The type of the aggregation table.

Details

The used aggregation functions are min,median,max,mean,sd, number and frequency.

Value

The function returns a data frame with a aggregation of specific columns.

analysePage *Collect metadata of webpage.*

Description

Analyse the Webpage and collect article attributes.

Usage

```
analysePage(mediaName = "fazBlog", loadedPage)
```

Arguments

mediaName	Name of digital media. Currently only Spiegel-Online, FAZ.net and Welt.de are supported.
loadedPage	Name of the website-file.

Value

The function extract article attributes from query result page or article page as data frame.

cleanCharacters *Remove false characters.*

Description

Removes special metadata characters from DBpedia-Results.

Usage

```
cleanCharacters(.df)
```

Arguments

.df	The data frame with results from DBpedia, preferably the output from function getConceptData.

Value

Data frame without metadata characters.

Examples

```
## Not run:
cleanCharacters(dat)

## End(Not run)
```

cleanData *Clean article data.*

Description

Cleans the article dataset.

Usage

```
cleanData(dat, mediaTarget)
```

Arguments

dat	Data frame with article data, preferably the output from function analysePage.
mediaTarget	Name of the media source (e.g. "spiegelOnline").

Details

This function removes unnecessary paragraphs and character from the article result set. It also sets the right data format (e.g. date) and encoding. In addition it parse interesting data from character vectors.

Value

The function returns a data frame with clean and correct formated results.

descrArticles *Describe article data.*

Description

Set of article descriptions.

Usage

```
descrArticles(.dat)
```

Arguments

.dat	The object with article data, preferably the output from function cleanData.

Details

The description-columns will be added to the input data frame.

Value

The funtion returns a data frame with article descriptions like "number of words".

dualPlot *Combination of linechart and barplot.*

Description

Creates a plot combination of a linechart and a barplot.

Usage

```
dualPlot(titleMain, titleSub = "", yAxisTitle = "", xAxisTitle = "",
  refText = "", dat, pdfName, y2AxisTitle = "", selectSearchTerm)
```

Arguments

titleMain	The main title (on top) of the plot.
titleSub	Sub-title (at bottom) of the plot.
yAxisTitle	A label for the y axis.
xAxisTitle	A label for the x axis.
refText	Name of the source or author of the plot.
dat	the The object with article data, preferably the output from function freqByTime.
pdfName	Character string naming a pdf-file for plot output
y2AxisTitle	A label for the second y axis.
selectSearchTerm	The search-term for which to create the plot (defining the relevent rows of dat).

Details

This is a function requiring a special data frame with specific columns. See also the function freqByTime which creates the required data structure.

Examples

```
## Not run:
dualPlot(
titleMain = "The main title",
titleSub = "Subtitle",
yAxisTitle = "Y-Axis",
xAxisTitle = "X-Axis",
refText = "Source:",
dat,
pdfName = "./data/plot/out.pdf",
y2AxisTitle = "Y2-Axis",
selectSearchTerm = "keyword")

## End(Not run)
```

---

filterResultSet *Remove false results.*

Description

Removes false entries from resultset.

Usage

```
filterResultSet(.df)
```

Arguments

.df	The data frame with results from DBpedia, preferably the output from function getConceptData.

Value

Data frame with removed invalid rows.

Examples

```
## Not run:
filterResultSet(dat)

## End(Not run)
```

freqByTime *Frequency by time*

Description

Shows the number of articles by time (year, month, date etc.)

Usage

```
freqByTime(dat, timeDim = "%Y", completeDim = TRUE)
```

Arguments

dat	The object with article data, preferably the output from function cleanData.
timeDim	Dimension of time in POSIX standard format.
completeDim	If set to true, the time dimension will be coninuous (e.g. a month with no articles will appear in the frequency table).

Value

Returns the frequency of elemtens and average number of words per article for the given time dimension (e.g. frequency by month) as data frame.

See Also

See also strptime for possible date conversions and formats.

Examples

```
## Not run:
freqByTime(dat,"%Y",TRUE)
freqByTime(dat,"%Y%m",TRUE)

## End(Not run)
```

getConceptData *Query DBpedia.*

Description

Queries the DBpedia using a SPARQL endpoint.

Usage

```
getConceptData(concept = "index", subjectLabel = NULL,
  spqlEndPoint = "http://de.dbpedia.org/sparql")
```

Arguments

concept	name of the Wikipedia concept where to look after corresponding subjects.
subjectLabel	if specified, additionally include matching subjects.
spqlEndPoint	the sparcle endpoint where the query is send to.

Value

The function returns all subjects from DBpedia which are associated to the given concept as a data frame.

---

getPages *get html*

Description

Get html-pages.

Usage

```
getPages(url, mediaName, searchTerm, pageCount)
```

Arguments

url	URL of the Webpage.
mediaName	Name of digital media. Currently only Spiegel-Online, FAZ.net and Welt.de are supported.
searchTerm	The keyword for which the search engine returned the URL.
pageCount	Number of resultpage from search engine.

Details

The name of the saved html-file will contain the name of the source media, the number of the resultpage the url was found and the searchterm used for getting the url.

Value

The function will download the webpage for a given url as html on harddisk in the project-subfolder.

getResultPages *Get query results.*

Description

Queries the search engine of a digital media site for a specific searchterm.

Usage

```
getResultPages(mediaName, searchTerm)
```

Arguments

mediaName	Name of digital media. Currently only Spiegel-Online, FAZ.net and Welt.de are supported.
searchTerm	The keyword the search engine should queried for.

Value

The function returns all result-pages of the queried website as html on harddisk in the project-subfolder.

noParagraph *Remove paragraphs from string.*

Description

Removes all paragraphs from a string variable. Besides apply also a left and right trim of spaces.

Usage

```
noParagraph(stringVar)
```

Arguments

stringVar	A character vector.

Value

Character vector with removed paragraphs.

removeHtml *Remove html-code from content*

Description

Removes html-code from a character content.

Usage

```
removeHtml(htmlContent)
```

Arguments

htmlContent	Character vector with html-code.

Details

This function removes unnecessary html-code from a character vector.

Value

Character vector with removed html-code.

Examples

```
## Not run:
removeHtml(htmlContent)

## End(Not run)
```

sampleArticles *Sample articles*

Description

Samples articles

Usage

```
sampleArticles(dat, minArt = 10, maxArt = 30, percent = 0.05)
```

Arguments

dat	The object with article data, preferably the output from function cleanData
minArt	Minimum of articles in sample.
maxArt	Maximum of articles in sample.
percent	Relative number of articles in sample.

Details

The default sampling rules are minimum 10 articles, maximum 30 articles and otherwise 5 percent of available articles.

Value

The function returns a data frame with sampled articles.

Examples

```
## Not run:
sampeArticles(dat,minArt=10,maxArt=30,percent=0.05)
sampeArticles(dat,minArt=10,maxArt=80,percent=0.20)

## End(Not run)
```

---

statTab *Arrange table.*

Description

Selection of columns.

Usage

```
statTab(.dat, type = "articleStatistic")
```

Arguments

.dat	The object with article data, preferably the output from function cleanData.
type	The type of the aggregation table.

Value

The function returns a data frame with a specific selection of columns.

Index