

Package ‘SocialMediaMineR’

February 19, 2015

Type Package

Title A Social Media Search and Analytic Tool

Version 0.1

Date 2014-06-25

Author Marco Toledo Bastos

Maintainer Marco Toledo Bastos <marco@toledobastos.com>

Imports RCurl, jsonlite, httr

Description SocialMediaMineR is a social media search and analytic tool that takes one or multiple URL(s) and returns the information about the popularity and reach of the URL(s) on social media. The function `get_socialmedia` retrieves the number of shares, likes, tweets, pins, and hits on Facebook, Twitter, Pinterest, StumbleUpon, LinkedIn, and Reddit. The package also includes dedicated functions for each social networking site and a function to decode shortened URLs.

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2014-06-25 14:48:55

R topics documented:

<code>get_facebook</code>	2
<code>get_linkedin</code>	3
<code>get_pinterest</code>	4
<code>get_reddit</code>	5
<code>get_socialmedia</code>	6
<code>get_stumbleupon</code>	7
<code>get_twitter</code>	8
<code>get_url</code>	9
<code>SocialMediaMineR</code>	10

Index	11
--------------	-----------

get_facebook	<i>Get the number of Facebook likes, shares, and comments of a specific URL</i>
--------------	---

Description

Takes a URL and returns a data frame with the original URL, normalized URL, number of Facebook shares, likes, comments, total hits, clicks, Facebook ID of comments, and comments box count. The function accepts the optional variable `sleep.time` (default=0) to avoid hitting API limits (if implemented). Use `sleep.time` to define the number of seconds the function will wait before proceeding to the next URL. If the URLs are shortened, you will have to first resolve them using the function `get_url`.

Usage

```
get_facebook(links, sleep.time = 0)
```

Arguments

<code>links</code>	URL to be tracked on Facebook
<code>sleep.time</code>	Number of seconds the function will wait before proceeding to the next URL (optional)

Value

Returns the original URL, normalized URL, number of Facebook shares, likes, comments, total hits, clicks, Facebook ID of comments, and comments box count.

Author(s)

Marco Toledo Bastos

References

Information about the Facebook API is available at <https://developers.facebook.com/>

Examples

```
get_facebook("http://www.google.com/")
```

get_linkedin	<i>Retrieves the share count of URLs on LinkedIn</i>
--------------	--

Description

Takes a URL and returns a data frame with the original URL and the number of shares on LinkedIn. The function accepts the optional variable `sleep.time` (default=0) to avoid hitting API limits (if implemented). Use `sleep.time` to define the number of seconds the function will wait before proceeding to the next URL. If the URLs are shortened, you will have to first resolve them using the function `get_url`.

Usage

```
get_linkedin(links, sleep.time = 0)
```

Arguments

<code>links</code>	URL to be tracked on LinkedIn
<code>sleep.time</code>	Number of seconds the function will wait before proceeding to the next URL (optional)

Value

Returns the original URL and the number of shares on LinkedIn.

Author(s)

Marco Toledo Bastos

References

Information about the LinkedIn API is available at <https://developer.linkedin.com/apis>

Examples

```
get_linkedin("http://www.nytimes.com/")
```

get_pinterest	<i>Takes a URL and retrieves the number of pins on Pinterest</i>
---------------	--

Description

Takes a URL and returns a data frame with the original URL and the number of pins on Pinterest. The function accepts the optional variable `sleep.time` (default=0) to avoid hitting API limits (if implemented). Use `sleep.time` to define the number of seconds the function will wait before proceeding to the next URL. If the URLs are shortened, you will have to first resolve them using the function `get_url`.

Usage

```
get_pinterest(links, sleep.time = 0)
```

Arguments

<code>links</code>	URL to be tracked on Pinterest
<code>sleep.time</code>	Number of seconds the function will wait before proceeding to the next URL

Value

Returns the original URL and the number of pins on Pinterest.

Author(s)

Marco Toledo Bastos

References

Pinterest is yet to release its official public API. More informatio at <https://developers.pinterest.com/>

Examples

```
get_pinterest("http://www.nytimes.com/")
```

`get_reddit`*Retrieves detailed information about URLs on Reddit*

Description

Takes a URL and returns a data frame with forty columns associated with the performance of the URL on Reddit. The columns are: domain, banned_by, media_embed, subreddit, selftext_html, selftext, likes, secure_media, link_flair_text, id, gilded, secure_media_embed, clicked, stickied, author, media, score, approved_by, over_18, hidden, thumbnail, subreddit_id, edited, link_flair_css_class, author_flair_css_class, downs, saved, is_self, permalink, name, created, url, author_flair_text, title, created_utc, ups, num_comments, visited, num_reports, and distinguished. The function accepts the optional variable `sleep.time` (default=0) to avoid hitting API limits (if implemented). Use `sleep.time` to define the number of seconds the function will wait before proceeding to the next URL. If the URLs are shortened, you will have to first resolve them using the function `get_url`.

Usage

```
get_reddit(links, sleep.time = 0)
```

Arguments

<code>links</code>	URL to be tracked on Reddit
<code>sleep.time</code>	Number of seconds the function will wait before proceeding to the next URL (optional)

Value

It returns a data frame with forty columns associated the performance of the URL on Reddit. The columns are: domain, banned_by, media_embed, subreddit, selftext_html, selftext, likes, secure_media, link_flair_text, id, gilded, secure_media_embed, clicked, stickied, author, media, score, approved_by, over_18, hidden, thumbnail, subreddit_id, edited, link_flair_css_class, author_flair_css_class, downs, saved, is_self, permalink, name, created, url, author_flair_text, title, created_utc, ups, num_comments, visited, num_reports, and distinguished.

Author(s)

Marco Toledo Bastos

References

The Reddit API documentation is available at <http://www.reddit.com/dev/api>

Examples

```
get_reddit("http://i.imgur.com/rSnA8.png")
```

get_socialmedia *A social media search and analytic tool*

Description

This is the main function of the package. It takes one or multiple URL(s) and returns the information about the popularity and reach of the URL(s) on several social media platforms. It retrieves the number of shares, likes, tweets, pins, and hits on Facebook, Twitter, Pinterest, StumbleUpon, LinkedIn, and Reddit. Please note it can be time consuming to search for a large number of URLs. This function also accepts the optional variable `sleep.time` (default=0) to avoid hitting API limits (if implemented). Use `sleep.time` to define the number of seconds the function will wait before proceeding to the next URL. If the URLs are shortened, you will have to first resolve them using the function `get_url`. Remember to include the trailing slash at the end of the address if you are searching for a domain.

Usage

```
get_socialmedia(links, sleep.time = 0)
```

Arguments

<code>links</code>	URL or list of URLs to be tracked across Facebook, Twitter, Pinterest, StumbleUpon, LinkedIn, and Reddit.
<code>sleep.time</code>	Number of seconds the function will wait before proceeding to the next URL (optional)

Details

Most social media APIs fail to return valid results for shortened URLs. For these cases you should first use the function `get_url` included in this the package that quickly resolves shortened URLs. The function `get_url` can be used together with `get_socialmedia`.

Value

Returns a data frame with the number of shares, likes, tweets, pins, and hits on Facebook, Twitter, Pinterest, StumbleUpon, LinkedIn, and Reddit. The resulting data frame includes the following 15 columns:

<code>url</code>	URL searched
<code>normalized_url</code>	URL returned
<code>fbk_shares</code>	number of Facebook shares
<code>fbk_likes</code>	number of Facebook likes
<code>fbk_comments</code>	number of Facebook comments
<code>fbk_total</code>	total number of hits on Facebook
<code>fbk_clicks</code>	number of Facebook clicks

twt_tweets	number of Twitter tweets
rdt_score	Reddit score
rdt_downs	number of downs on Reddit
rdt_ups	number of ups on Reddit
rdt_comments	number of comments on Reddit
lkn_shares	number of LinkedIn shares
stu_views	number of StumbleUpon views
pin_counts	number of pins on Pinterest

Author(s)

Marco Toledo Bastos

Examples

```
get_socialmedia("http://cran.r-project.org/")
```

get_stumbleupon	<i>Retrieves the view count of URLs on StumbleUpon</i>
-----------------	--

Description

Takes a URL and returns a data frame with the original URL and the view count on StumbleUpon. The function accepts the optional variable `sleep.time` (default=0) to avoid hitting API limits (if implemented). Use `sleep.time` to define the number of seconds the function will wait before proceeding to the next URL. If the URLs are shortened, you will have to first resolve them using the function `get_url`.

Usage

```
get_stumbleupon(links, sleep.time = 0)
```

Arguments

<code>links</code>	URL to be tracked on StumbleUpon
<code>sleep.time</code>	Number of seconds the function will wait before proceeding to the next URL (optional)

Value

Returns a data frame with the view count of URLs on StumbleUpon. The resulting data frame includes the following 12 columns: `result.url`, `result.in_index`, `result.publicid`, `result.views`, `result.title`, `result.thumbnail`, `result.thumbnail_b`, `result.submit_link`, `result.badge_link`, `result.info_link`, `timestamp`, and `success`.

Author(s)

Marco Toledo Bastos

References

StumbleUpon has an API for the short URL functionality offered by Su.pr. More information at <http://www.stumbleupon.com/help/business-tools/supr/supr-api/>

Examples

```
get_stumbleupon("http://www.cnn.com")
```

get_twitter

Retrieves the tweet count of URLs on Twitter

Description

Takes a URL and returns a data frame with the original URL and the tweet count on Twitter. The function accepts the optional variable `sleep.time` (default=0) to avoid hitting API limits (if implemented). Use `sleep.time` to define the number of seconds the function will wait before proceeding to the next URL. If the URLs are shortened, you will have to first resolve them using the function `get_url`.

Usage

```
get_twitter(links, sleep.time = 0)
```

Arguments

<code>links</code>	URL to be tracked on Twitter
<code>sleep.time</code>	Number of seconds the function will wait before proceeding to the next URL (optional)

Value

Returns the original URL and the number of shares on Twitter.

Author(s)

Marco Toledo Bastos

References

Information about the Twitter API is available at <https://dev.twitter.com/>

Examples

```
get_twitter("http://www.cnn.com")
```

`get_url`*Unshortens any URL created using URL shortening services.*

Description

The function takes a character vector with shortened URLs and returns a character vector (unless `return.df=T`) with the resolved URLs. The original shortened URLs are stored in the returned vector as the `names` attribute of the vector. The function `get_url` accepts the optional variables `sleep.time` (default=0) and `return.df` (default=FALSE). The variable `sleep.time` is used to avoid hitting API limit and defines the number of seconds the function will wait before proceeding to the next URL. The default value of `sleep.time` is zero. The variable `return.df` is used to return a data frame instead of a character vector. The default value of `return.df` is FALSE.

Usage

```
get_url(links, sleep.time = 0, return.df = F)
```

Arguments

<code>links</code>	Shortened URL or list of URLs to be resolved
<code>sleep.time</code>	Number of seconds the function will wait before proceeding to the next URL (optional)
<code>return.df</code>	If <code>return.df=TRUE</code> the function returns a data frame. If <code>return.df=FALSE</code> (default) the function returns a character vector

Value

If `return.df=TRUE` the function returns a data frame with two columns: the original URL and the resolved URL. If `return.df=FALSE` (default) the function returns a character vector with the resolved URL.

Author(s)

Marco Toledo Bastos

References

The function relies on Hadley Wickham's `httr` package. Hadley was also very helpful in optimizing the code.

Examples

```
get_url("http://t.co/iWmeUiECAa")
```

Description

Search and analytic tool that takes one or multiple URL(s) and returns the information about the popularity and reach of the URL(s) on social media. It retrieves the number of shares, likes, tweets, pins, and hits on Facebook, Twitter, Pinterest, StumbleUpon, LinkedIn, and Reddit.

Details

Package: SocialMediaMineR
Type: Package
Version: 0.1
Date: 2014-06-20
License: GPL (>= 2)

The function `get_socialmedia` returns a data frame with the number of shares, likes, tweets, pins, and hits on Facebook, Twitter, Pinterest, StumbleUpon, LinkedIn, and Reddit. The resulting data frame includes the following 15 columns:

url URL searched normalized_url URL returned fbk_shares number of Facebook shares fbk_likes number of Facebook likes fbk_comments number of Facebook comments fbk_total total number of hits on Facebook fbk_clicks number of Facebook clicks twt_tweets number of tweets on Twitter rdt_score Reddit score rdt_downs number of downs on Reddit rdt_ups number of ups on Reddit rdt_comments number of comments on Reddit lkn_shares number of LinkedIn shares stu_views number of StumbleUpon views pin_counts number of pins on Pinterest

The package also includes dedicated functions to each social networking site. The function `get_facebook` performs a search on Facebook, `get_linkedin` on LinkedIn, `get_pinterest` on Pinterest, `get_reddit` on Reddit, `get_stumbleupon` on StumbleUpon, and `get_twitter` on Twitter. The function `get_url` resolves shortened URLs and can be used with the aforementioned functions.

Author(s)

Marco Toledo Bastos

Maintainer: Marco Toledo Bastos <marco@toledobastos.com>

Examples

```
get_socialmedia("http://www.nytimes.com/")
```

Index

*Topic **IO**

- get_facebook, 2
- get_linkedin, 3
- get_pinterest, 4
- get_socialmedia, 6
- get_stumbleupon, 7
- get_twitter, 8
- get_url, 9
- SocialMediaMineR, 10

*Topic **OI**

- get_reddit, 5

*Topic **utilits**

- get_facebook, 2
- get_linkedin, 3
- get_pinterest, 4
- get_reddit, 5
- get_socialmedia, 6
- get_stumbleupon, 7
- get_twitter, 8
- get_url, 9

- get_facebook, 2
- get_linkedin, 3
- get_pinterest, 4
- get_reddit, 5
- get_socialmedia, 6
- get_stumbleupon, 7
- get_twitter, 8
- get_url, 9

- SocialMediaMineR, 10
- SocialMediaMineR-package
(SocialMediaMineR), 10