Package ‘scientoText’

October 14, 2022

Type Package
Title Text & Scientometric Analytics
Version 0.1
Date 2016-07-20
Author Ashraf Uddin
Maintainer Ashraf Uddin <mdaakib18@gmail.com>
Description It involves bibliometric indicators calculation from bibliometric data. It also deals pattern analysis using the text part of bibliometric data. The bibliometric data are obtained from mainly Web of Science and Scopus.
License GPL-3
LazyData TRUE
Imports stringr, tm, utils
RoxygenNote 5.0.1
NeedsCompilation no
Repository CRAN
Date/Publication 2016-07-20 23:55:39

R topics documented:

authorship_pattern ................................................. 2
author_info ......................................................... 2
citation_info ....................................................... 3
country_pattern .................................................... 4
g_index ............................................................. 5
highly_cited ....................................................... 5
h_index ............................................................. 6
international_col .................................................. 7
international_colmat ............................................... 8
term_freq ........................................................ 9

Index 10

1
**authorship_pattern**

*Co-authorship Matrix and Average co-authorship*

**Description**

It finds year-wise co-authorship matrix and average co-authorship values

**Usage**

```r
authorship_pattern(authors, pub_years, sep)
```

**Arguments**

- `authors`: A character vector containing author names
- `pub_years`: A numeric vector containing publication years
- `sep`: A character pattern separating author names

**Value**

A list with co-authorship matrix and average co-authorship values

**Examples**

```r
authorship_pattern(authors,years,"\"",")
```

---

**author_info**

*Authors' Information*

**Description**

It finds top author names and their different performance indicators

**Usage**

```r
author_info(authors, citations, sep, top = 10, only_first_author = F)
```

**Arguments**

- `authors`: A character vector containing author names
- `citations`: A numeric vector containing citations
- `sep`: A character pattern separating author names
- `top`: The number of top authors
- `only_first_author`: Logical. If to find the author list by the first authors
citation_info

Value

A list consisting of author names, total instances, total citations, h index, g index, i10 index, max citation

See Also

g index h index

Examples

authors<-c("Wolf W.R., Lele S.K.",
"Shin D., Yeh X., Khatib O.",
"Aukes D., Heyneman B., Duchaine V., Cutkosky M.R.")
author_info(authors,c(3,4,1),',','

country_pattern

Country Instances

Description
Country-wise and year-wise output for a defined period.

Usage

country_pattern(affiliations, pub_years = NULL, countries = NULL, only_first_author = F)

Arguments

affiliations A text vector containing affiliation (country) information
pub_years A numeric vector containing publication years
countries A list of countries (optional)
only_first_author Logical. If to find the author list by the first authors

Details
The function returns year and country-wise output matrix if the publication years are provided. If only affiliation data is provided the country-wise output is returned as a single vector instead of a matrix.

Value
A list containing country output and other details.

Examples

affiliations<-c("Stanford University, Stanford, CA, United States; Montreal, QC, Canada", "Stanford University, United States; Google Inc., United States", "University of Michigan, Ann Arbor, MI 48109-2122, United States; Tsinghua University, Beijing 100084, China", "Imperial College London, London, SW7 2BZ, United Kingdom; ENSTA, Ecole Polytechnique, Palaiseau, 91761, France")


country_pattern( affiliations, pub_years)
country_pattern(affiliations)
**g_index**

<table>
<thead>
<tr>
<th>g_index</th>
<th>g index</th>
</tr>
</thead>
</table>

**Description**

g index

**Usage**

```r
g_index(citations)
```

**Arguments**

| citations | A numeric vector containing citations |

**Value**

return the g index for the given citations

**See Also**

h index

**Examples**

```r
g_index(c(1,2,5,0,3,11))
```

---

**highly_cited**

*Highly Cited Instances*

<table>
<thead>
<tr>
<th>highly_cited</th>
<th>Highly Cited Instances</th>
</tr>
</thead>
</table>

**Description**

It finds the number of highly cited instances year-wise.

**Usage**

```r
highly_cited(citations, pub_years, ref_citations = NULL,
              ref_pub_years = NULL, top = NULL, year_lim = list())
```

**Arguments**

<table>
<thead>
<tr>
<th>citations</th>
<th>A numeric vector containing citations</th>
</tr>
</thead>
<tbody>
<tr>
<td>pub_years</td>
<td>A numeric vector containing publication years</td>
</tr>
<tr>
<td>ref_citations</td>
<td>The citations of reference instances</td>
</tr>
<tr>
<td>ref_pub_years</td>
<td>The publication years of reference instances</td>
</tr>
<tr>
<td>top</td>
<td>An integer which defines top percent highly cited instances</td>
</tr>
<tr>
<td>year_lim</td>
<td>A list conating years and year-wise citation threshold. If not mentioned these values are calculated from ref_citations, ref_pub_years &amp; top.</td>
</tr>
</tbody>
</table>
Value

Returns a list containing number of top highly cited instances with other details

Examples

citations<-c(2,0,12,3,1,4,5,8,2)
ref_citations<-c(3,0,12,3,1,41,5,8,2,2,0,12,30,1,1,4,5,8,12)
highly_cited(citations,pub_years,ref_citations,ref_pub_years,10)
highly_cited(citations,pub_years,year_lim = list(c(2011, 2012, 2013), c(41, 12, 12)))

h_index

Description

Find h index for a given set of documents

Usage

h_index(citations)

Arguments

citations A numeric vector containing citations

Value

return the h index for the given citations

References

Hirsch, J. E. (2005). An index to quantify an individual’s scientific research output. Proceedings of
the National academy of Sciences of the United States of America, 102(46), 16569-16572.

See Also

g_index

Examples

h_index(c(1,2,5,0,3,11))
international_col

**Description**

Calculate the number of Internationally Collaborated Papers

**Usage**

```r
ingernational_col(affiliations, pub_years = NULL, countries = NULL)
```

**Arguments**

- `affiliations`: A text vector containing affiliation (country) information
- `pub_years`: A numeric vector containing publication years
- `countries`: A list of countries (optional)

**Details**

It finds if there is any International Collaboration so affiliation fields must have country information

**Value**

Collaboration count or a list (collaboration counts year-wise)

**Examples**

```r
affiliations<-c("Stanford University, Stanford, CA, United States; Montreal, QC, Canada",
                 "Stanford University, United States; Google Inc., United States",
                 "University of Michigan, Ann Arbor, MI 48109-2122, United States;
                 Tsinghua University, Beijing 100084, China",
                 "Imperial College London, London, SW7 2BZ, United Kingdom;
                 ENSTA, Ecole Polytechnique, Palaiseau, 91761, France")


international_col( affiliations, pub_years)
international_col(affiliations)
```
international_colmat   International Collaboration Matrix

Description

Calculate Internationally Collaborated Matrix(es)

Usage

international_colmat(affiliations, pub_years = NULL, countries = NULL)

Arguments

affiliations   A text vector containing affiliation (country) information
pub_years     A numeric vector containing publication years
countries      A list of countries (optional)

Details

It finds the collaboration network at international level in terms of adjacent matrix so affiliation fields must have country information

Value

Collaboration adjacent matrix(es)

Examples

affiliations<-c("Stanford University, Stanford, CA, United States; Montreal, QC, Canada","Stanford University, United States; Google Inc., United States","University of Michigan, Ann Arbor, MI 48109-2122, United States; Tsinghua University, Beijing 100084, China","Imperial College London, London, SW7 2BZ, United Kingdom; ENSTA, Ecole Polytechnique, Palaiseau, 91761, France")

international_colmat( affiliations, pub_years)
rinternational_colmat(affiliations)
**term_freq**

---

**Term Frequency**

**Description**

Term Frequency

**Usage**

```r
term_freq(text, pub_years = NULL, sep = NULL, top = NULL)
```

**Arguments**

- `text`: A character vector
- `pub_years`: A numeric vector containing publication years
- `sep`: A character value which separates the terms (optional)
- `top`: The number of terms to return

**Value**

Term frequency vector or matrix (for year-wise)
Index

author_info, 2
authorship_pattern, 2

citation_info, 3
country_pattern, 4

g_index, 5, 6

h_index, 6
highly_cited, 5

international_col, 7
international_colmat, 8

term_freq, 9