Package 'bibliometrix'

October 18, 2016

Type Package

Title Bibliometric and Co-Citation Analysis Tool

Version 1.2

Date 2016-09-07

Description Tool for quantitative research in scientometrics and bibliometrics.
It provides various routines for importing bibliographic data from SCOPUS and Thomson Reuters' ISI Web of Knowledge databases, performing bibliometric analysis and building data matrices for co-citation, coupling and scientific collaboration analysis.

License GPL-3

URL http://www.bibliometrix.org

LazyData FALSE

Encoding latin1

Depends R (>= 3.3.0)

Imports stats, dplyr, Matrix, stringr

Suggests igraph, FactoMineR, factoextra, knitr, rmarkdown

RoxygenNote 5.0.1

NeedsCompilation no

Author Massimo Aria [cre, aut],
Corrado Cuccurullo [aut]

Maintainer Massimo Aria <aria@unina.it>

VignetteBuilder knitr

Repository CRAN

Date/Publication 2016-10-18 10:20:45

R topics documented:

`bibliometrix-package` *Tool for quantitative research in scientometrics and bibliometrics.*

Description

It provides various routines for importing bibliographic data from SCOPUS and Thomson Reuters' ISI Web of Knowledge databases, perfoming bibliometric analysis and building data matrices for co-citation, coupling and scientific collaboration analysis.

Details

Package:	bibliometrix
Type:	Package
Version:	0.1
Date:	2016-05-05
License:	GPL-3

Author(s)

Massimo Aria <massimo.aria@unina.it>, Corrado Cuccurullo <corrado.cuccurullo@unina2.it>

Maintainer: Massimo Aria <massimo.aria@unina.it>

References

Cuccurullo C., Aria M., Sarto F. (2016) Foundations and Trends in Performance Management. A Twenty-five Years Bibliometric Analysis in Business and Public Administration Domains, Scientometrics, DOI: 10.1007/s11192-016-1948-8.

Koseoglu, M. A. (2016). Growth and structure of authorship and co-authorship network in the strategic management realm: Evidence from the Strategic Management Journal. BRQ Business Research Quarterly.

Batagelj, V., Cerinsek, M. (2013). On bibliographic networks. Scientometrics, 96(3), 845-864.

Yan, E., Ding, Y. (2012). Scholarly network similarities: How bibliographic coupling networks, citation networks, cocitation networks, topical networks, coauthorship networks, and coword networks relate to each other. Journal of the American Society for Information Science and Technology, 63(7), 1313-1326.

Rousseau, R. (2010). Bibliographic coupling and co-citation as dual notions. The Janus faced scholar. A Festschrift in honour of Peter Ingwersen, 173-183.

Leydesdorff, L., Vaughan, L. (2006). Co-occurrence matrices and their applications in information science: Extending ACA to the Web environment. Journal of the American Society for Information Science and technology, 57(12), 1616-1628.

Examples

```
## load scientometrics data set
# data(scientometrics_text)

## Convert text data into a bibliographic data frame
# scient_df <- convert2df(scientometrics_text, dbsource="isi", format="plaintext")

## Perform a bibliometric analysis of the bibliographic data frame
# results <- biblioAnalysis(scient_df)

## summarize results
# summary(results, k=10, pause=FALSE)

## plot results
# plot(results, k=10, pause=FALSE)

## Estimate Lotka's law coefficients
# L=lotka(results)
# L

## Perform authors' dominance analysis
#DF=dominance(results)
```

```
#DF
```

`biblio` *Dataset of "Bibliometrics" manuscripts.*

Description

The set of manuscripts which the title containing the word "bibliometrics" and published in a journal indexed by ISI WoK database.
Period: 2006 - 2015
Database: ISI Web of Knowledge

Usage

```
biblio
```

Format

A large character with 9014 rows.
Data has been imported by an ISI Export file in bibtex format using the function `readLines`.

Source

http://www.webofknowledge.com

`biblioAnalysis` *Bibliometric Analysis*

Description

It performs a bibliometric analysis of a dataset imported from SCOPUS and Thomson Reuters' ISI Web of Knowledge databases.

Usage

```
biblioAnalysis(M, sep = ";")
```

Arguments

`M`	is a bibliographic data frame obtained by the converting function `convert2df`. It is a data matrix with cases corresponding to manuscripts and variables to Field Tag in the original SCOPUS and Thomson Reuters' ISI Web of Knowledge file.
`sep`	is the field separator character. This character separates strings in each column of the data frame. The default is `sep = ";"`.

Value

`biblioAnalysis` returns an object of `class` "bibliometrix".

The functions `summary` and `plot` are used to obtain or print a summary and some useful plots of the results.

An object of `class` "bibliometrix" is a list containing the following components:

Articles	the total number of manuscripts
Authors	the authors' frequency distribution
AuthorsFrac	the authors' frequency distribution (fractionalized)
FirstAuthors	first author of each manuscript
nAUperPaper	the number of authors per manuscript
Apparences	the number of author apparences
nAuthors	the number of authors
AuMultiAuthoredArt	the number of authors of multi authored articles
Years	pubblication year of each manuscript
FirstAffiliation	the affiliation of the first author
Affiliations	the frequency distribution of affiliations (of all co-authors for each paper)
Aff_frac	the fractionalized frequency distribution of affiliations (of all co-authors for each paper)
CO	the affiliation country of first author
Countries	the affiliation countries' frequency distribution
TotalCitation	the number of times each manuscript has been cited
TCperYear	the yearly average number of times each manuscript has been cited
Sources	the frequency distribution of sources (journals, books, etc.)
DE	the frequency distribution of authors' keywords
ID	the frequency distribution of keywords associated to the manuscript bySCOPUS and Thomson Reu

See Also

`convert2df` to import and convert an ISI or SCOPUS Export file in a bibliographic data frame.

`summary` to obtain a summary of the results.

`plot` to draw some useful plots of the results.

Examples

```
data(scientometrics)

results <- biblioAnalysis(scientometrics)

summary(results, k = 10, pause = FALSE)
```

`biblioNetwork` *Creating Bibliographic networks*

Description

`biblioNetwork` creates different bibliographic networks from a bibliographic data frame.

Usage

```
biblioNetwork(M, analysis = "coupling", network = "authors", sep = ";")
```

Arguments

M	is a bibliographic data frame obtained by the converting function convert2df. It is a data matrix with cases corresponding to manuscripts and variables to Field Tag in the original SCOPUS and Thomson Reuters' ISI Web of Knowledge file.
analysis	is a character object. It indicates the type of analysis have to be performed. analysis argument can be "collaboration", "coupling" or "co-citation". Default is analysis = "coupling".
network	is a character object. It indicates the network typology. The network aurgument can be "authors", "references", "sources", "countries","keywords" or "author_keywords". Default is network = "authors".
sep	is the field separator character. This character separates strings in each column of the data frame. The default is sep = ";".

Details

The function biblioNetwork can create a collection of bibliographic networks following the approach proposed by Batagely and Cerinsek (2013).

Typical networks output of biblioNetwork are:

Collaboration Networks
– Authors collaboration (analysis = "collaboration", network = "authors")
– Country collabortion (analysis = "collaboration", network = "countries")

Co-citation Networks
– Authors co-citation (analysis = "co-citation", network = "authors")
– Reference co-citation (analysis = "co-citation", network = "references")
– Source co-citation (analysis = "co-citation", network = "sources")

Coupling Networks
– Manuscript coupling (analysis = "coupling", network = "references")
– Authors coupling (analysis = "coupling", network = "authors")
– Source coupling (analysis = "coupling", network = "sources")
– Keyword coupling (analysis = "coupling", network = "keywords")
– Author-Keyword coupling (analysis = "coupling", network = "author_keywords")
– Country coupling (analysis = "coupling", network = "countries")

Value

It is a squared network matrix. It is an object of class dgMatrix of the package Matrix.

See Also

convert2df to import and convert a SCOPUS and Thomson Reuters' ISI Web of Knowledge export file in a data frame.

cocMatrix to compute a co-occurrence matrix.

biblioAnalysis to perform a bibliometric analysis.

Examples

```
# EXAMPLE 1: Authors collaboration network

library(igraph)
data(scientometrics)

NetMatrix <- biblioNetwork(scientometrics, analysis = "collaboration",
network = "authors", sep = ";")
netDegree <- 2
diag <- Matrix::diag
NetMatrix <- NetMatrix[diag(NetMatrix) >= netDegree,diag(NetMatrix) >= netDegree]
diag(NetMatrix) <- 0

bsk.network <- graph.adjacency(NetMatrix,mode = "undirected")
plot(bsk.network,layout = layout.fruchterman.reingold, vertex.label.dist = 0.5,
vertex.frame.color = 'blue', vertex.label.color = 'black',
vertex.label.font = 1, vertex.label = V(bsk.network)$name, vertex.label.cex = 0.7)


# EXAMPLE 2: Co-citation network

library(igraph)
data(scientometrics)

NetMatrix <- biblioNetwork(scientometrics, analysis = "co-citation",
network = "references", sep = ";")
netDegree=10
diag <- Matrix::diag
NetMatrix <- NetMatrix[diag(NetMatrix) >= netDegree,diag(NetMatrix) >= netDegree]
diag(NetMatrix) <- 0

bsk.network <- graph.adjacency(NetMatrix,mode = "undirected")
plot(bsk.network,layout = layout.fruchterman.reingold, vertex.label.dist = 0.5,
vertex.frame.color = 'blue', vertex.label.color = 'black',
vertex.label.font = 1, vertex.label = V(bsk.network)$name, vertex.label.cex = 0.7)
```

---

biblio_df *Dataset of "Bibliometrics" manuscripts.*

Description

The set of manuscripts which the title containing the word "bibliometrics" and published in a journal indexed by ISI WoK database.
Period: 2006 - 2015
Database: ISI Web of Knowledge

Usage

```
biblio_df
```

Format

#' A data frame with 99 rows (manuscripts) and 16 variables (ISI tag field):

AU Authors

TI Document Title

SO Publication Name (or Source)

JI ISO Source Abbreviation

DT Document Type

DE Author Keywords

ID Keywords associated by ISI or SCOPUS database

AB Abstract

C1 Author Address

RP Reprint Address

CR Cited References

TC Times Cited

PY Year

SC Subject Category

UT Unique Article Identifier

DB Database

Source

http://www.webofknowledge.com

`citations` *Citation frequency distribution*

Description

It calculates frequency distribution of citations.

Usage

```
citations(M, field = "article", sep = ";")
```

Arguments

`M`	is a bibliographic data frame obtained by the converting function `convert2df`. It is a data matrix with cases corresponding to manuscripts and variables to Field Tag in the original SCOPUS and Thomson Reuters' ISI Web of Knowledge file.
`field`	is a character. It can be "article" or "author" to obtain frequency distribution of cited citations or cited authors (only first authors for ISI database) respectively. The default is `field = "article"`.
`sep`	is the field separator character. This character separates citations in each string of CR column of the bibiographic data frame. The default is `sep = ";"`.

Value

an object of `class` "list" containing the following components:

Cited	the most frequent cited manuscripts or authors
Year	the pubblication year (only for cited article analysis)
Source	the journal (only for cited article analysis)

See Also

`biblioAnalysis` function for bibliometric analysis.

`summary` to obtain a summary of the results.

`plot` to draw some useful plots of the results.

Examples

```
## EXAMPLE 1: Cited articles

data(scientometrics)

CR <- citations(scientometrics, field = "article", sep = ";")

CR$Cited[1:10]
CR$Year[1:10]
CR$Source[1:10]
```

```
## EXAMPLE 2: Cited first authors

data(scientometrics)

CR <- citations(scientometrics, field = "author", sep = ";")

CR$Cited[1:10]
```

---

cocMatrix *Co-occurrence matrix*

Description

cocMatrix computes co-occurences between elements of a Tag Field from a bibliographic data frame. Manuscript is the unit of analysis.

Usage

```
cocMatrix(M, Field = "AU", type = "sparse", sep = ";")
```

Arguments

M	is a data frame obtained by the converting function convert2df. It is a data matrix with cases corresponding to articles and variables to Field Tag in the original ISI or SCOPUS file.
Field	is a character object. It indicates one of the field tags of the standard ISI WoS Field Tag codify. Field can be equal to one of this tags:

AU	Authors
SO	Publication Name (or Source)
JI	ISO Source Abbreviation
DE	Author Keywords
ID	Keywords associated by ISI or SCOPUS database
CR	Cited References

for a complete list of filed tags see: ISI WoS Field Tags

type — indicates the output format of co-occurrences:

type = "matrix"	produces an object of class matrix
type = "sparse"	produces an object of class dgMatrix of the package Matrix. "sparse" argument generates a compact

sep — is the field separator character. This character separates strings in each column of the data frame. The default is sep = ";".

Details

This co-occurrence matrix can be tranformed into a collection of compatible networks. Through matrix multiplication you can obtain different networks. The fuction follows the approach proposed by Batagely and Cerinsek (2013).

Value

a co-occurrence matrix with cases corresponding to manuscripts and variables to the objects extracted from the Tag `Field`.

See Also

`convert2df` to import and convert an ISI or SCOPUS Export file in a data frame.

`biblioAnalysis` to perform a bibliometric analysis.

`biblioNetwork` to compute a bibliographic network.

Examples

```
# EXAMPLE 1: Articles x Authors co-occurrence matrix

data(scientometrics)
WA <- cocMatrix(scientometrics, Field = "AU", type = "sparse", sep = ";")

# EXAMPLE 2: Articles x Cited References co-occurrence matrix

# data(scientometrics)

# WCR <- cocMatrix(scientometrics, Field = "CR", type = "sparse", sep = ";")

# EXAMPLE 3: Articles x Cited First Authors co-occurrence matrix

# data(scientometrics)
# scientometrics <- metaTagExtraction(scientometrics, Field = "CR_AU", sep = ";")
# WCR <- cocMatrix(scientometrics, Field = "CR_AU", type = "sparse", sep = ";")
```

---

convert2df *Convert an ISI or SCOPUS Export file into a data frame*

Description

It converts a SCOPUS and Thomson Reuters' ISI Web of Knowledge export file and create a data frame from it, with cases corresponding to articles and variables to Field Tag in the original file.

Usage

```
convert2df(file, dbsource = "isi", format = "bibtex")
```

Arguments

file	is a character array containing data read from an ISI WoK Export file (in plain text or bibtex format) or SCOPUS Export file (exclusively in bibtex format).
dbsource	is a character indicating the bibliographic database. dbsource can be "isi" or "scopus". Default is dbsource = "isi".
format	is a character indicating the format of the SCOPUS and Thomson Reuters' ISI Web of Knowledge export file. format can be "bibtex" or "plaintext". Default is format = "bibtex".

Details

Actually the function allows to convert both SCOPUS/ISI files in bibtext format and just ISI files in plain text format.

Value

a data frame with cases corresponding to articles and variables to Field Tag in the original export file.

data frame columns are named using the standard ISI WoS Field Tag codify. The main field tags are:

AU	Authors
TI	Document Title
SO	Publication Name (or Source)
JI	ISO Source Abbreviation
DT	Document Type
DE	Authors' Keywords
ID	Keywords associated by SCOPUS or ISI database
AB	Abstract
C1	Author Address
RP	Reprint Address
CR	Cited References
TC	Times Cited
PY	Year
SC	Subject Category
UT	Unique Article Identifier
DB	Database

for a complete list of filed tags see: ISI WoS Field Tags

See Also

scopus2df for converting SCOPUS Export file (in bibtex format)

isibib2df for converting ISI Export file (in bibtex format)

isi2df for converting ISI Export file (in plain text format)

Other converting.functions: isi2df, isibib2df, scopus2df

Examples

```
# An ISI or SCOPUS Export file can be read using \code{\link{readLines}} function:

# largechar <- readLines('filename.txt')

# filename.txt is an ISI or SCOPUS Export file in plain text or bibtex format.
# The file have to be saved without Byte order mark (U+FEFF) at the beginning
# and EoF code at the end of file.
# The original file (exported by ISI or SCOPUS search web site) can be modified
# using an advanced text editor like Notepad++ or Emacs.

#  biblio <- readLines('http://www.bibliometrix.org/datasets/bibliometrics_articles.txt')

data(biblio)

biblio_df_df <- convert2df(file = biblio, dbsource = "isi", format = "bibtex")
```

---

countries *Index of Countries.*

Description

Data frame containing a normalized index of countries.
Data are used by biblioAnalysis function to extract Country Field of Cited References and Authors.

Usage

```
countries
```

Format

A data frame with 198 rows and 1 variable:

countries country names

couplingSimilarity *Coupling similarity index*

Description

It calculates a relative measure of bibliographic coupling.

Usage

```
couplingSimilarity(NetMatrix, type = "jaccard")
```

Arguments

NetMatrix	is a coupling matrix obtained by the network functions biblioNetwork or cocMatrix.
type	is a character. It can be "jaccard" or "salton" to obtain Jaccard or Salton similarity index respectively. The default is type = "jaccard".

Details

couplingSimilarity calculates Jaccard or Salton similarity from a coupling bibliographic matrix.

The Jaccard index (or Jaccard similarity coefficient) gives us a relative measure of the overlap of two sets. It is calcultated as the ratio between the intersection and the union of the reference lists (of two manuscripts). The Salton index, instead, relates the intersection of the two lists to the geometric mean of the size of both sets.

Both indices are equal to zero if the intersection of the reference lists is empty; its reach a maximum of one if both lists are identical.

Value

a similarity matrix.

See Also

biblioNetwork function to compute a bibliographic network.

cocMatrix to compute a bibliographic bipartite network.

Examples

```
data(scientometrics)
NetMatrix <- biblioNetwork(scientometrics, analysis = "coupling", network = "references", sep = ";")
S=couplingSimilarity(NetMatrix, type = "jaccard")
```

dfMerge *Merging of Bibliographic data frames*

Description

Merge two bibliographic data frames.

Usage

```
dfMerge(M1, M2, Field = "TI", tol = 0.9)
```

Arguments

M1	is the first bibliographic data frame.
M2	is the second bibliographic data frame.
Field	is a character object. It indicates one of the field tags used to match the two data frames. Field can be equal to one of this tags: TI (title), AB (abstract), UT (manuscript ID).
tol	is a numeric value giving the minimum relative similarity to marge two manuscripts. Default value is `tol = 0.90`.

Details

A bibliographic data frame is obtained by the converting function `convert2df`. It is a data matrix with cases corresponding to manuscripts and variables to Field Tag in the original SCOPUS and Thomson Reuters' ISI Web of Knowledge file. The function merges two bibliographic data frames deleting duplicate manuscripts. Duplicate entries are identified through the generalized Levenshtein (edit) distance. Two manuscripts that have a similarity greater than `tol` argument are stored in the merged data frame only once.

Value

the value returned from `dfMerge` is a data frame containing the merged data frames.

See Also

`convert2df` to import and convert an ISI or SCOPUS Export file in a bibliographic data frame.

`biblioAnalysis` function for bibliometric analysis.

`summary` to obtain a summary of the results.

`plot` to draw some useful plots of the results.

Examples

```
data(scientometrics)

M1=scientometrics[1:20,]

M2=scientometrics[10:30,]

mergedM <- dfMerge(M1, M2, Field = "TI", tol = 0.95)

dim(mergedM)
```

`dominance` *Authors' dominance ranking*

Description

It calculates the authors' dominance ranking from an object of the class 'bibliometrix' as proposed by Kumar & Kumar, 2008.

Usage

```
dominance(results, k = 10)
```

Arguments

`results`	is an object of the class 'bibliometrix' for which the analysis of the authors' dominance ranking is desired.
`k`	is an integer, used for table formatting (number of authors). Default value is 10.

Value

The function `dominance` returns a data frame with cases corresponding to the first `k` most productive authors and variables to typical field of a dominance analysis.

the data frame variables are:

`Dominance Factor`	Dominance Factor (DF = FAA / MAA)
`Multi Authored`	N. of Multi Authored Articles (MAA)
`First Authored`	N. of First Authored Articles (FAA)
`Rank by Articles`	Author Ranking by N. of Articles
`Rank by DF`	Author Ranking by Dominance Factor

See Also

`biblioAnalysis` function for bibliometric analysis

summary method for class 'bibliometrix'

Examples

```
data(scientometrics)
results <- biblioAnalysis(scientometrics)
DF=dominance(results)
DF
```

---

garfield *Eugene Garfield's manuscripts.*

Description

All manuscripts published by Eugene Garfield.
Period: 1954 - 2014
Database: SCOPUS source

Usage

```
garfield
```

Format

A data frame with 147 rows and 15 variables:

AU Authors
TI Document Title
SO Publication Name (or Source)
JI ISO Source Abbreviation
DT Document Type
DE Author Keywords
ID Keywords associated by ISI or SCOPUS database
AB Abstract
C1 Author Address
RP Reprint Address
CR Cited References
TC Times Cited
PY Year
UT Unique Article Identifier
DB Database

Source

http://www.scopus.com

Hindex *h-index calculation*

Description

It calculates the authors' h-index and its variants.

Usage

```
Hindex(M, authors, sep = ";")
```

Arguments

M	is a bibliographic data frame obtained by the converting function convert2df. It is a data matrix with cases corresponding to manuscripts and variables to Field Tag in the original SCOPUS and Thomson Reuters' ISI Web of Knowledge file.
authors	is a character vector. It contains the the authors' names list for which you want to calculate the H-index. The aurgument has the form C("SURNAME1 N","SURNAME2 N",...), in other words, for each author: surname and initials separated by one blank space. i.e for the auhtors SEMPRONIO TIZIO CAIO and ARIA MASSIMO authors argument is authors = c("SEMPRONIO TC", "ARIA M").
sep	is the field separator character. This character separates auhtors in each string of AU column of the bibiographic data frame. The default is sep = ";".

Value

an object of class "list". It contains two elements: H is a data frame with h-index, g-index and m-index for each author; CitationList is a list with the bibliographic collection for each author.

See Also

convert2df to import and convert an ISI or SCOPUS Export file in a bibliographic data frame.

biblioAnalysis function for bibliometric analysis.

summary to obtain a summary of the results.

plot to draw some useful plots of the results.

Examples

```
### EXAMPLE 1: ###

data(scientometrics)

authors <- c("SMALL H", "CHEN DZ")

Hindex(scientometrics, authors, sep = ";")$H
```

```
### EXAMPLE 2: Garfield h-index###

data(garfield)

indices=Hindex(garfield, authors="GARFIELD E", sep = ";")

# h-index, g-index and m-index of Eugene Garfield
indices$H

# Papers and total citations
indices$CitationList[[1]]
```

---

histNetwork *Historical co-citation network*

Description

histNetwork creates a historical co-citation network from a bibliographic data frame.

Usage

```
histNetwork(M, n = 10, sep = ";")
```

Arguments

M	is a bibliographic data frame obtained by the converting function convert2df. It is a data matrix with cases corresponding to manuscripts and variables to Field Tag in the original SCOPUS and Thomson Reuters' ISI Web of Knowledge file.
n	is an integer, indicating the number of most cited references to select. Default value is 10.
sep	is the field separator character. This character separates strings in CR column of the data frame. The default is sep = ";".

Value

histAnalysis returns an object of class "list" containing the following components:

NetMatrix	the historical co-citation network matrix
Degree	the min degree of the network
histData	the set of n most cited references

See Also

convert2df to import and convert an ISI or SCOPUS Export file in a bibliographic data frame.

summary to obtain a summary of the results.

plot to draw some useful plots of the results.

biblioNetwork to compute a bibliographic network.

Examples

```
data(scientometrics)

histResults <- histNetwork(scientometrics, n = 10, sep = ";")

library("igraph")

# Create igraph object
bsk.network <- graph.adjacency(histResults[[1]],mode = "directed")

# Remove loops
bsk.network <- simplify(bsk.network, remove.multiple = TRUE, remove.loops = TRUE)

# Create the network layout (fixing vertical vertex coordinates by years)
l <- layout.fruchterman.reingold(bsk.network)
l[,2] <- histResults[[3]]$Year

# Plot the hystorical co-citation network
plot(bsk.network,layout = l, vertex.label.dist = 0.5,
vertex.frame.color = 'blue', vertex.label.color = 'black',
vertex.label.font = 1, vertex.label = V(bsk.network)$name,
vertex.label.cex = 0.5,edge.arrow.size=0.1)
```

---

isi2df *Convert an ISI WoK Export file into a data frame*

Description

It converts an ISI Wok Export file and create a data frame from it, with cases corresponding to articles and variables to Field Tag in the original file.

Usage

```
isi2df(D)
```

Arguments

D	is a character array containing data read from a ISI Export file (in plain text format).

Value

a data frame with cases corresponding to articles and variables to Field Tag in the original ISI file.

See Also

scopus2df for converting SCOPUS Export file (in bibtex format)

Other converting.functions: convert2df, isibib2df, scopus2df

Examples

```
# An ISI Export file can be read using \code{\link{readLines}} function:

# largechar <- readLines('filename.txt')

# filename.txt is an ISI Export file in plain text format.
# The file have to be saved without Byte order mark (U+FEFF) at the beginning
# and EoF code at the end of file.
# The original file (exported by ISI search web site) can be modified
# using an advanced text editor like Notepad++ or Emacs.

# scientometrics_text <- readLines('http://www.bibliometrix.org/datasets/scientometrics.txt')
# data(scientometrics_text)
# scient_df <- isi2df(scientometrics_text)
```

---

isibib2df *Convert an ISI WoK Export file into a data frame*

Description

It converts an ISI WoK Export file and create a data frame from it, with cases corresponding to articles and variables to Field Tag in the original file.

Usage

```
isibib2df(D)
```

Arguments

D	is a character array containing data read from an ISI Export file (in bibtex format).

Value

a data frame with cases corresponding to articles and variables to Field Tag in the original SCOPUS file.

See Also

isi2df for converting ISI Export file (in plain text format)

Other converting.functions: convert2df, isi2df, scopus2df

Examples

```
# A ISI Export file can be read using \code{\link{readLines}} function:

# largechar <- readLines('filename.bib')

# filename.bib is an ISI Export file in plain text format.
# The file have to be saved without Byte order mark (U+FEFF) at the
# beginning and EoF code at the end of file.
# The original file (exported by ISI search web site) can be modified
# using an advanced text editor like Notepad++ or Emacs.

# largechar <- readLines('http://www.bibliometrix.org/datasets/ranking.bib')

# ranking <- isibib2df(largechar)
```

---

keywordAssoc *ID and DE keyword associations*

Description

It associates authors' keywords to keywords plus.

Usage

```
keywordAssoc(M, sep = ";", n = 10, excludeKW = NA)
```

Arguments

M	is a bibliographic data frame obtained by the converting function convert2df. It is a data matrix with cases corresponding to manuscripts and variables to Field Tag in the original SCOPUS and Thomson Reuters' ISI Web of Knowledge file.
sep	is the field separator character. This character separates keywords in each string of ID and DE columns of the bibiographic data frame. The default is sep = ";".
n	is a integer. It indicates the number of authors' keywords to associate to each keyword plus. The default is n = 10.
excludeKW	is character vector. It contains authors' keywords to exclude from the analysis.

Value

an object of class "list".

See Also

convert2df to import and convert an ISI or SCOPUS Export file in a bibliographic data frame.

biblioAnalysis function for bibliometric analysis.

summary to obtain a summary of the results.

plot to draw some useful plots of the results.

Examples

```
data(scientometrics)

KWlist <- keywordAssoc(scientometrics, sep = ";",n = 10, excludeKW = NA)

# list of first 10 Keywords plus
names(KWlist)

# list of first 10 authors' keywords associated to the first Keyword plus
KWlist[[1]][1:10]
```

`KeywordGrowth` *Yearly cumulative occurrences of top keywords/terms*

Description

It calculates yearly cumulative occurrences of top keywords/terms.

Usage

```
KeywordGrowth(M, Tag = "ID", sep = ";", top = 10)
```

Arguments

`M`	is a data frame obtained by the converting function `convert2df`. It is a data matrix with cases corresponding to articles and variables to Field Tag in the original ISI or SCOPUS file.
`Tag`	is a character object. It indicates one of the keyword field tags of the standard ISI WoS Field Tag codify (ID or DE) or a field tag created by `termExtraction` function (TI_TM, AB_TM, etc.).
`sep`	is the field separator character. This character separates strings in each keyword column of the data frame. The default is `sep = ";"`.
`top`	is a numeric. It indicates the number of top keywords to analize. The default value is 10.

Value

an object of class `data.frame`

Examples

```
data(scientometrics)
topKW=KeywordGrowth(scientometrics, Tag = "ID", sep = ";", top=5)
topKW

# Plotting results
#
# library(reshape2)
# library(ggplot2)
# DF=melt(topKW, id='Year')
# ggplot(DF,aes(Year,value, group=variable, color=variable))+geom_line()
```

`localCitations` *Author local citations*

Description

It calculates frequency distribution of cited local authors.

Usage

```
localCitations(M, results, sep = ";")
```

Arguments

`M`	is a bibliographic data frame obtained by the converting function `convert2df`. It is a data matrix with cases corresponding to manuscripts and variables to Field Tag in the original SCOPUS and Thomson Reuters' ISI Web of Knowledge file.
`results`	is an object of `class` "bibliometrix". The default is `field = "article"`.
`sep`	is the field separator character. This character separates citations in each string of CR column of the bibiographic data frame. The default is `sep = ";"`.

Details

Local citations measure how many times an author included in this collection have been cited by other authors also in the collection.

Value

an object of `class` "table".

See Also

`citations` function for citation frequency distribution.

`biblioAnalysis` function for bibliometric analysis.

`summary` to obtain a summary of the results.

`plot` to draw some useful plots of the results.

Examples

```
data(scientometrics)

results <- biblioAnalysis(scientometrics)

CR <- localCitations(scientometrics, results, sep = ";")

CR[1:10]
```

---

lotka *Lotka's law coefficient estimation*

Description

It estimates Lotka's law coefficients for scientific productivity (*Lotka A.J., 1926*)

Usage

```
lotka(results)
```

Arguments

results	is an object of the class 'bibliometrix' for which the analysis of the authors' dominance ranking is desired.

Value

The function lotka returns a list of summary statistics of the Lotka's law estimation of an object of class bibliometrix.

the list contains the following objects:

Beta	Beta coefficient
C	Constant coefficient
R2	Goodness of Fit
fitted	Fitted Values
p.value	Pvalue of two-sample Kolmogorov-Smirnov test between the empirical and the theorical Lotka's Law distril
AuthorProd	Authors' Productivity frequency table

See Also

biblioAnalysis function for bibliometric analysis

summary method for class 'bibliometrix'

Examples

```
data(scientometrics)
```

```
results <- biblioAnalysis(scientometrics)
L=lotka(results)
L
```

`metaTagExtraction` *Meta-Field Tag Extraction*

Description

It extracts some new useful field tags from standard ISI/SCOPUS fied tag codify.

Usage

```
metaTagExtraction(M, Field = "CR_AU", sep = ";")
```

Arguments

`M`	is a data frame obtained by the converting function `convert2df`. It is a data matrix with cases corresponding to articles and variables to Field Tag in the original ISI or SCOPUS file.
`Field`	is a character object. New tag exctracted from aggregated data is specified by this string. Field can be equal to one of this tags:

`"CR_AU"`	First Author of each cited reference
`"CR_SO"`	Source of each cited reference
`"AU_CO"`	Country of affiliation for each co-author
`"AU_UN"`	University of affiliation for each co-author

`sep`	is the field separator character. This character separates strings in each column of the data frame. The default is `sep = ";"`.

Value

the bibliometric data frame with a new column containing data about new field tag indicated in the argument `Field`.

See Also

`scopus2df` for converting ISO or SCPUS Export file into a data frame.

`biblioAnalysis` function for bibliometric analysis

Examples

```
# Example 1: First Authors for each cited reference

data(scientometrics)
```

```
scientometrics <- metaTagExtraction(scientometrics, Field = "CR_AU", sep = ";")
unlist(strsplit(scientometrics$CR_AU[1], ";"))


#Example 2: Source for each cited reference

data(scientometrics)
scientometrics <- metaTagExtraction(scientometrics, Field = "CR_SO", sep = ";")
unlist(strsplit(scientometrics$CR_SO[1], ";"))

#Example 3: Affiliation country for co-author

data(scientometrics)
scientometrics <- metaTagExtraction(scientometrics, Field = "AU_CO", sep = ";")
scientometrics$AU_CO[1:10]
```

`plot.bibliometrix` *Plotting bibliometric analysis results*

Description

`plot` method for class 'bibliometrix'

Usage

```
## S3 method for class 'bibliometrix'
plot(x, ...)
```

Arguments

`x`	is the object for which plots are desired.
`...`	can accept two arguments: `k` is an integer, used for plot formatting (number of objects). Default value is 10. `pause` is a logical, used to allow pause in screen scrolling of results. Default value is `pause = TRUE`.

Value

none. The function `plot` returns a set of plots of the object of class `bibliometrix`.

See Also

The bibliometric analysis function `biblioAnalysis`.

`summary` to compute a list of summary statistics of the object of class `bibliometrix`.

Examples

```
data(scientometrics)

results <- biblioAnalysis(scientometrics)

plot(results, k = 10, pause = FALSE)
```

`scientometrics` *"Co-citation analysis" and "Coupling analysis" manuscripts.*

Description

Manuscripts about the topics "co-citation analysis" and "coupling analysis" published on Scientometrics Journal.
Period: 1985 - 2015
Database: ISI Web of Knowledge

Usage

```
scientometrics
```

Format

A data frame with 147 rows and 16 variables:

AU Authors

TI Document Title

SO Publication Name (or Source)

JI ISO Source Abbreviation

DT Document Type

DE Author Keywords

ID Keywords associated by ISI or SCOPUS database

AB Abstract

C1 Author Address

RP Reprint Address

CR Cited References

TC Times Cited

PY Year

SC Subject Category

UT Unique Article Identifier

DB Database

Source

http://www.webofknowledge.com

---

scientometrics_text *"Co-citation analysis" and "Coupling analysis" manuscripts.*

Description

Manuscripts about the topics "co-citation analysis" and "coupling analysis" published on Scientometrics Journal.
Period: 1985 - 2015
Database: ISI Web of Knowledge

Usage

```
scientometrics_text
```

Format

A large character with 12731 rows.
Data has been imported by an ISI Export file in plain text format using the function readLines.

Source

http://www.webofknowledge.com

---

scopus2df *Convert a SCOPUS Export file into a data frame*

Description

It converts a SCOPUS Export file and create a data frame from it, with cases corresponding to articles and variables to Field Tag in the original file.

Usage

```
scopus2df(D)
```

Arguments

D	is a character array containing data read from a SCOPUS Export file (in bibtex format).

Value

a data frame with cases corresponding to articles and variables to Field Tag in the original SCOPUS file.

See Also

isi2df for converting ISI Export file (in plain text format)

Other converting.functions: convert2df, isi2df, isibib2df

Examples

```
# A SCOPUS Export file can be read using \code{\link{readLines}} function:

# largechar <- readLines('filename.bib')

# filename.bib is a SCOPUS Export file in plain text format.
# The file have to be saved without Byte order mark (U+FEFF) at the
# beginning and EoF code at the end of file.
# The original file (exported by SCOPUS search web site) can be modified
# using an advanced text editor like Notepad++ or Emacs.

#largechar <- readLines('http://www.bibliometrix.org/datasets/scopus.bib')


#scopus_df <- scopus2df(largechar)
```

---

stopwords *List of English stopwords.*

Description

A character vector containing a complete list of English stopwords
Data are used by biblioAnalysis function to extract Country Field of Cited References and Authors.

Usage

```
stopwords
```

Format

A character vector with 665 rows.

summary.bibliometrix *Summarizing bibliometric analysis results*

Description

summary method for class 'bibliometrix'

Usage

```
## S3 method for class 'bibliometrix'
summary(object, ...)
```

Arguments

object	is the object for which a summary is desired.
...	can accept two arguments: k integer, used for table formatting (number of rows). Default value is 10. pause locical, used to allow pause in screen scrolling of results. Default value is pause = TRUE.

Value

The function summary computes and returns a list of summary statistics of the object of class bibliometrics.

the list contains the following objects:

MainInformation	Main Information about Data
AnnualProduction	Annual Scientific Production
AnnualGrowthRate	Annual Percentage Growth Rate
MostProdAuthors	Most Productive Authors
MostProdCountries	Most Productive Countries
TCperCountries	Total Citation per Countries
MostRelSources	Most Relevant Sources
MostRelKeywords	Most Relevant Keywords

See Also

biblioAnalysis function for bibliometric analysis

plot to draw some useful plots of the results.

Examples

```
data(scientometrics)

results <- biblioAnalysis(scientometrics)

summary(results)
```

tableTag *Tabulate elements from a Tag Field column*

Description

It tabulates elements from a Tag Field column of a bibliographic data frame.

Usage

```
tableTag(M, Tag = "CR", sep = ";")
```

Arguments

M	is a data frame obtained by the converting function convert2df. It is a data matrix with cases corresponding to articles and variables to Field Tag in the original ISI or SCOPUS file.
Tag	is a character object. It indicates one of the field tags of the standard ISI WoS Field Tag codify.
sep	is the field separator character. This character separates strings in each column of the data frame. The default is sep = ";".

Details

tableTag is an internal routine of main function biblioAnalysis.

Value

an object of class table

Examples

```
data(scientometrics)
Tab <- tableTag(scientometrics, Tag = "CR", sep = ";")
Tab[1:10]
```

termExtraction *Term extraction tool from textual fields of a manuscript*

Description

It extracts terms from a textual field (abstract, title, author's keywords, etc.) of a bibliographic data frame.

Usage

```
termExtraction(M, Field = "TI", remove.numbers = TRUE,
  remove.terms = NULL, keep.terms = NULL, verbose = TRUE)
```

Arguments

M	is a data frame obtained by the converting function convert2df. It is a data matrix with cases corresponding to articles and variables to Field Tag in the original ISI or SCOPUS file.
Field	is a character object. It indicates the field tag of textual data : "TI" Manuscript title "AB" Manuscript abstract "ID" Manuscript author's keywords The default is Field = "TI".
remove.numbers	is logical. If TRUE all numbers are deleted from the documents before term extraction. The default is remove.numbers = TRUE.
remove.terms	is a character vector. It contains a list of additional terms to delete from the documents before term extraction. The default is remove.terms = NULL.
keep.terms	is a character vector. It contains a list of compound words "formed by two or more terms" to keep in their original form in the term extraction process. The default is keep.terms = NULL.
verbose	is logical. If TRUE the function prints the most frequent terms extracted from documents. The default is verbose=TRUE.

Value

the bibliometric data frame with a new column containing terms about the field tag indicated in the argument Field.

See Also

convert2df to import and convert an ISI or SCOPUS Export file in a bibliographic data frame.

biblioAnalysis function for bibliometric analysis

Examples

```
# Example 1: Term extraction from titles

data(scientometrics)

# vector of compound words
keep.terms=c("co-citation analysis","bibliographic coupling")

# term extraction
scientometrics <- termExtraction(scientometrics, Field = "TI",
remove.numbers=TRUE, remove.terms=NULL, keep.terms=keep.terms, verbose=TRUE)

# terms extracted from the first 10 titles
scientometrics$TI_TM[1:10]


#Example 2: Term extraction from abstracts

data(scientometrics)

# vector of terms to remove
remove.terms=c("analysis","bibliographic")

# term extraction
scientometrics <- termExtraction(scientometrics, Field = "AB",
 remove.numbers=TRUE, remove.terms=remove.terms, keep.terms=NULL, verbose=TRUE)

# terms extracted from the first abstract
scientometrics$AB_TM[1]
```

`timeslice` *Bibliographic data frame time slice*

Description

Divide a bibliographic data frame into time slice

Usage

```
timeslice(M, breaks = NA, k = 5)
```

Arguments

`M`	is a bibliographic data frame obtained by the converting function `convert2df`. It is a data matrix with cases corresponding to manuscripts and variables to Field Tag in the original SCOPUS and Thomson Reuters' ISI Web of Knowledge file.
`breaks`	is a numeric vector of two or more unique cut points.

k — is a integer value giving the number of intervals into which the data frame is to be cut. k is used only in case breaks argument is not provided. The default is k = 5.

Value

the value returned from split is a list containing the data frames for each sub-period.

See Also

convert2df to import and convert an ISI or SCOPUS Export file in a bibliographic data frame.

biblioAnalysis function for bibliometric analysis.

summary to obtain a summary of the results.

plot to draw some useful plots of the results.

Examples

```
data(scientometrics)

list_df <- timeslice(scientometrics, breaks = c(1995, 2005))

names(list_df)
```

---

trim.leading *Deleting leading white spaces*

Description

Deleting leading white spaces from a character object.

Usage

```
trim.leading(x)
```

Arguments

x — is a character object.

Details

tableTag is an internal routine of bibliometrics package.

Value

an object of class character

Examples

```
char <- c("  Alfred", "Mary", " John")
char
trim.leading(char)
```

Index