

Package: doilinker (via r-universe)

August 24, 2024

Title Link Preprints And Publications By DOI

Version 0.1.1

Date 2024-02-26

Description Links preprints to publications using the method described in Cabanac G, Oikonomidi T, Boutron I. ``Day-to-day discovery of preprint-publication links". *Scientometrics*. 2021;1–20. DOI: 10.1007/s11192-021-03900-7.

License GPL (>= 3)

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.3

Imports purrr, rcrossref, SnowballC, stopwords, stringi, stringr

Suggests rmarkdown, knitr, testthat (>= 3.0.0), spelling, covr

Config/testthat/edition 3

VignetteBuilder knitr

Language en-GB

URL <https://github.com/lazappi/doilinker>

BugReports <https://github.com/lazappi/doilinker/issues>

Repository <https://lazappi.r-universe.dev>

RemoteUrl <https://github.com/lazappi/doilinker>

RemoteRef HEAD

RemoteSha 16a860d1ed696c4cdf8d10ba3a75e09482097a3

Contents

calc_jaccard_similarity	2
get_cr_authors_str	2
get_first_orcid	3

is_doi_preprint	3
is_result_match	4
same_first_author	5
search_doi_links	5

Index	7
--------------	----------

calc_jaccard_similarity
Calculate Jaccard similarity

Description

Calculate the Jaccard similarity between two strings. Strings are first tokenised, stop words are removed and tokens are stemmed.

Usage

```
calc_jaccard_similarity(string1, string2)
```

Arguments

string1	First string to score similarity
string2	Second string to score similarity

Value

Jaccard similarity score

get_cr_authors_str *Get Crossref authors string*

Description

Convert the authors table from Crossref into a single string. It has the form "LastName1, FirstName1; LastName2, FirstName2;...".

Usage

```
get_cr_authors_str(cr_authors)
```

Arguments

cr_authors	Authors data.frame from a Crossref query
------------	--

Value

Character vector with authors names

get_first_orcid	<i>Get first author ORCID</i>
-----------------	-------------------------------

Description

Return the ORCID for the first author of a reference.

Usage

```
get_first_orcid(cr_authors)
```

Arguments

cr_authors Authors data.frame from a Crossref query

Value

Character vector with ORCID or NA if not found

is_doi_preprint	<i>Is DOI preprint</i>
-----------------	------------------------

Description

Check whether a DOI is from a preprint server.

Usage

```
is_doi_preprint(doi, container_title)
```

Arguments

doi DOI to check
container_title Whether or not there is a container title (journal name) associated with the DOI

Value

Logical whether the DOI is from a preprint

<code>is_result_match</code>	<i>Is result a match</i>
------------------------------	--------------------------

Description

Decide if a results from a DOI query is a match or not.

Usage

```
is_result_match(  
  query_title,  
  result_title,  
  query_orcid,  
  result_orcid,  
  query_authors,  
  result_authors  
)
```

Arguments

<code>query_title</code>	Title of the query reference
<code>result_title</code>	Title of the result reference
<code>query_orcid</code>	ORCID of the query first author
<code>result_orcid</code>	ORCID of the result first author
<code>query_authors</code>	Authors string for the query
<code>result_authors</code>	Authors string for the result

Details

Result is a match if the similarity between titles is sufficiently high or if the similarity is lower by the first author ORCIDs match or the first author names match.

Value

Logical whether the result is a match or not

same_first_author	<i>Same first author</i>
-------------------	--------------------------

Description

Check if two author strings have the same first author.

Usage

```
same_first_author(authors_str1, authors_str2)
```

Arguments

authors_str1 Character vector containing first authors list
authors_str2 Character vector containing second authors list

Details

Last names of the first authors are compared. If these match then first names are checked and TRUE is returned if either the whole first name or first initial matches. If first and last names cannot be separated the whole author names are compared. Some simplification of characters in names is performed to improve matches.

Value

Logical whether first authors are the same

search_doi_links	<i>Search DOI links</i>
------------------	-------------------------

Description

Search Crossref for potential linked references for a DOI

Usage

```
search_doi_links(  
  doi,  
  preprint = NULL,  
  limit = 20,  
  filter_matches = FALSE,  
  verbose = TRUE  
)
```

Arguments

doi	DOI to query
preprint	Logical. Whether doi is a preprint (TRUE) or publication (FALSE). If NULL (default) attempt to determine this from doi.
limit	Number of Crossref results to return. Default is 20, max is 1000.
filter_matches	Logical. If TRUE only return matches rather than all Crossref results.
verbose	Logical. Whether or not to print progress messages.

Details

Based on the method described in "Day-to-day discovery of preprint–publication links" <https://doi.org/10.1007/s11192-021-03900-7> and code from <https://github.com/gcabanac/preprint-publication-linker>. Query filters have been modified to allow reverse linking publications to preprints as well as preprints to publications.

Value

tibble of potential links

Examples

```
search_doi_links("10.1101/133173")
```

Index

`calc_jaccard_similarity`, 2

`get_cr_authors_str`, 2

`get_first_orcid`, 3

`is_doi_preprint`, 3

`is_result_match`, 4

`same_first_author`, 5

`search_doi_links`, 5