

Package: geoparser (via r-universe)

September 17, 2024

Type Package

Title Detect Country Names in Documents

Version 0.1

Description Detects country names in PDF documents imported with the package 'pdftools'.

URL <https://github.com/frbcesab/geoparser>

BugReports <https://github.com/frbcesab/geoparser/issues>

License GPL (>= 2)

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.3

Depends R (>= 2.10)

Imports stringr, tidy

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition 3

VignetteBuilder knitr

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

RemoteUrl <https://github.com/FRBCesab/geoparser>

RemoteRef HEAD

RemoteSha 1da32855198145af687ffd76a58f34d4d461abe8

Contents

geoparser	2
world_countries	3
Index	4

geoparser

Detect countries names

Description

Detect countries names

Usage

```
geoparser(x)
```

Arguments

x a character in which countries will be detected.

Value

A data.frame with the following four columns:

- geographic_entity: the name of the country
- n_pages: the total number of pages in the document
- page: the page number
- count: the occurrence of the country name for a given page

Examples

```
## Example document ----
texte <- c(
  "
  Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod
  tempor incididunt US labore et dolore magna aliqua. USA enim ad minim
  veniam, quis nostrud exercitation ullamco laboris United States
  ",
  "
  Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod
  tempor incididunt US labore et dolore magna aliqua. USA enim ad minim
  veniam, quis nostrud exercitation ullamco laboris Canada.
  "
)

## Detect countries ----
geoparser(texte)
```

world_countries	<i>Dataset: World countries list</i>
-----------------	--------------------------------------

Description

This dataset contains a list of 256 World countries (geographic entities) according to the GADM database <https://gadm.org>.

Usage

```
world_countries
```

Format

A `data.frame` with 256 rows (geographic entities, official (or non official) countries) and the 10 following variables:

iso_alpha2 the ISO 3166 alpha 2 code of the country

iso_alpha3 the ISO 3166 alpha 3 code of the country

iso_num the ISO 3166 numeric code of the country

geographic_entity the name of the country (geographic entity)

sovereignty the name of recognized country (by the UN)

formal_name the formal name of the geographic entity

parser the regular expression used to detect the entity in a full text

continent the name of the continent where the entity is located

un_region the name of the UN region where the entity is located

un_subregion the name of the UN subregion where the entity is located

Examples

```
data("world_countries")  
head(world_countries)
```

Index

* **datasets**

world_countries, 3

geoparser, 2

world_countries, 3