CHAS_report

Provided by the Dutch Cultural Heritage Agency (test version)

Date created: 2023-09-28 14:16:50

Cultural Heritage ASessor: Collection Report

This is a template to connect a linked data version of a collection to different thesauri and term lists. The examples here can be replaced with other queries, as long as the thesaurus or list provides a SPARQL endpoint.

Information on how to use this template can be found in the README of the CHAS repository.

Load the required R packages

```r
# Check if packages are installed on machine

packages_to_install <- c("jsonlite", "rjson", "data.table", "httr", "curl",
                         "stringi", "rmarkdown")

is_package_installed <- function(package_name) {
  return(package_name %in% row.names(installed.packages()))
}

# Loop through the list of packages
for (package in packages_to_install) {
  if (!is_package_installed(package)) {
    # If the package is not installed, install it
    install.packages(package, dependencies = TRUE)
  }
}

# Load required packages
library(jsonlite)
library(rjson)
library(data.table)
library(httr)
library(curl)
library(stringi)
library(rmarkdown)
```

Overview of terms in collection

```r
url <- "https://api.linkeddata.cultureelerfgoed.nl/datasets/ruben-schalk/colonial/services/europeana-col
sparql_query <- "PREFIX dct: <http://purl.org/dc/terms/>
PREFIX schema: <https://schema.org/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

          SELECT ?obj (count(?obj) as ?n) WHERE {

                ?sub dct:medium|dct:subject|schema:artform ?obj .


              }
ORDER BY DESC (?n)
LIMIT 20"


response <- POST(url, body = list(query = sparql_query))
content <- content(response, "parsed")
df <- as.data.frame(do.call(rbind, lapply(content, as.data.frame)))
print(df)
```

```
##                 obj  n
## 1           textiel 63
## 2   Oosters tapijt 60
## 3            tapijt 60
## 4    wol (textiel) 60
## 5    beeldhouwkunst 50
## 6          geknoopt 49
## 7         sculptuur 46
## 8    figuurplastiek 35
## 9       miscellanea 35
## 10        aardewerk 31
## 11     archeologica 30
## 12         fragment 30
## 13          pigment 30
## 14           zilver 25
## 15            hout 22
## 16           katoen 22
## 17           kralen 21
## 18          geweven 20
## 19            zijde 17
## 20             tas 14
```

Connecting to the Cultural Heritage Thesaurus

```r
url <- "https://api.linkeddata.cultureelerfgoed.nl/datasets/ruben-schalk/colonial/services/europeana-co
sparql_query <- "PREFIX dct: <http://purl.org/dc/terms/>
PREFIX schema: <https://schema.org/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

                SELECT (count(distinct(?obj)) as ?terms) (count(distinct(?uri_cht)) as ?cht_matches) WH

                   ?sub dct:medium|dct:subject|schema:artform ?obj .

                   BIND(strlang(?obj, 'nl') AS ?obj_nl)

              SERVICE <https://api.linkeddata.cultureelerfgoed.nl/datasets/rce/Cultuurhistorische-Thes

                   GRAPH <https://data.cultureelerfgoed.nl/term/id/cht/thesaurus> {

                      optional{   ?uri_cht skos:prefLabel|skos:altLabel|skos:hiddenLabel ?obj_nl . }

    }}
                }
    LIMIT 10"


response <- POST(url, body = list(query = sparql_query))
content <- content(response, "parsed")
df <- as.data.frame(do.call(rbind, lapply(content, as.data.frame)))
print(df)
```

```
##   terms cht_matches
## 1   180         122
```

Connecting to the Colonial History Thesaurus

```r
url <- "https://api.linkeddata.cultureelerfgoed.nl/datasets/ruben-schalk/colonial/services/europeana-co
sparql_query <- "PREFIX dct: <http://purl.org/dc/terms/>
                 PREFIX schema: <https://schema.org/>
                 PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
                 PREFIX dc: <http://purl.org/dc/elements/1.1/>
                 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
                 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

                 SELECT (count(distinct(?obj)) as ?terms) (count(distinct(?uri_koloniaal)) as ?koloniaal

                    ?sub dct:medium|dct:subject|schema:artform ?obj .

                    BIND(strlang(?obj, 'nl') AS ?obj_nl)

                 SERVICE <https://api.linkeddata.cultureelerfgoed.nl/datasets/rce/Koloniaal-Verleden/serv
```

```
                optional{   ?uri_koloniaal skos:prefLabel|skos:altLabel|skos:hiddenLabel ?obj_n
        }
            } LIMIT 10"


response <- POST(url, body = list(query = sparql_query))
content <- content(response, "parsed")
df <- as.data.frame(do.call(rbind, lapply(content, as.data.frame)))
print(df)
```

```
##   terms koloniaal_matches
## 1   180                 9
```

Connecting to Geonames

```
url <- "https://api.linkeddata.cultureelerfgoed.nl/queries/ruben-schalk/colonial-geonames/run"
sparql_query <- "PREFIX dct: <http://purl.org/dc/terms/>
PREFIX schema: <https://schema.org/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX skosxl: <http://www.w3.org/2008/05/skos-xl#>
PREFIX aat: <http://vocab.getty.edu/aat/>
PREFIX gvp: <http://vocab.getty.edu/ontology#>
PREFIX gn: <https://www.geonames.org/ontology#>

SELECT (count(distinct(?obj)) as ?colonial_spatial) (count(distinct(?uri)) as ?geonames_match) WHERE {
  ?sub dct:spatial ?obj .

  SERVICE <https://demo.netwerkdigitaalerfgoed.nl/geonames/sparql> {
    optional { ?uri gn:name|gn:alternateName ?obj }
  }
}
LIMIT 10"

response <- POST(url, body = list(query = sparql_query))
content <- content(response, "parsed")
df <- as.data.frame(do.call(rbind, lapply(content, as.data.frame)))
print(df)
```

```
##   colonial_spatial geonames_match
## 1               55             10
```

Connecting descriptions to the Contested Terms Knowledge Graph

```r
### Import description-field as text from collection dataset

url <- "https://api.linkeddata.cultureelerfgoed.nl/datasets/ruben-schalk/rce-art-sample/services/rce-ar
query <- "PREFIX dct: <http://purl.org/dc/terms/>
          PREFIX schema: <https://schema.org/>
          PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
          PREFIX dc: <http://purl.org/dc/elements/1.1/>
          PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
          PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

          SELECT ?identifier ?description WHERE {

           ?identifier dct:title ?description .


          }
          LIMIT 500"

response <- POST(url, body = list(query = query))
content <- content(response, "parsed")
colonial <- as.data.frame(do.call(rbind, lapply(content, as.data.frame)))

### Import terms from word matter knowledge graph ###

url_wordsmatter <- "https://api.linkeddata.cultureelerfgoed.nl/datasets/rce/Words-Matter-Knowledge-Grap
query <- "PREFIX dct: <http://purl.org/dc/terms/>
          PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
          PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
          SELECT DISTINCT ?obj WHERE {
             ?sub <http://www.w3.org/2008/05/skos-xl#literalForm> ?obj .

             FILTER(LANG(?obj) = 'nl')
             } "

response <- POST(url_wordsmatter, body = list(query = query))
content <- content(response, "parsed")
wordsm <- as.data.frame(do.call(rbind, lapply(content, as.data.frame)))

### Find terms from description in collection in words matter graph ###


# Function to check if any word from word_df appears in the text_variable of text_df

check_words_in_text <- function(word_df, text_df, word_column, text_variable) {
  # Create an empty logical vector to store results
  match_results <- logical(nrow(text_df))

  # Create an empty list to store the matched words for each row
  matched_words_list <- vector("list", length = nrow(text_df))

  # Loop through the words in word_df and check if they appear in any row of text_df's text_variable
  for (i in seq_len(nrow(text_df))) {
    text <- text_df[[text_variable]][i]
```

```
    words <- stri_extract_all_regex(text, paste0("\\b", word_df[[word_column]], "\\b"), opts_regex = li
    matched_words <- unique(na.omit(unlist(words)))
    if (length(matched_words) > 0) {
      match_results[i] <- TRUE
      matched_words_list[[i]] <- matched_words
    }
  }

  # Convert the list of matched words to a character vector
  matched_words <- sapply(matched_words_list, function(x) {
    if (length(x) == 0) return("")
    paste(x, collapse = ", ")
  })

  return(list(match_results, matched_words))
}

# Call the function to find matches between 'obj' in wordsm and 'description' in colonial
results <- check_words_in_text(wordsm, colonial, "obj", "description")

# Extract the match_results and matched_words from the results list
matches <- results[[1]]
matched_words <- results[[2]]

# Add the match_results and matched_words as new columns to colonial
colonial$matches <- matches
colonial$matched_words <- matched_words
setDT(colonial)
print(colonial[matched_words != "", ])
```

```
##                                                             identifier
## 1: https://linkeddata.cultureelerfgoed.nl/cho-kennis/rce-art-sample/id/rce_131257
## 2: https://linkeddata.cultureelerfgoed.nl/cho-kennis/rce-art-sample/id/rce_131258
## 3: https://linkeddata.cultureelerfgoed.nl/cho-kennis/rce-art-sample/id/rce_131260
## 4: https://linkeddata.cultureelerfgoed.nl/cho-kennis/rce-art-sample/id/rce_131264
## 5: https://linkeddata.cultureelerfgoed.nl/cho-kennis/rce-art-sample/id/rce_133248
## 6: https://linkeddata.cultureelerfgoed.nl/cho-kennis/rce-art-sample/id/rce_133250
## 7: https://linkeddata.cultureelerfgoed.nl/cho-kennis/rce-art-sample/id/rce_133251
## 8:  https://linkeddata.cultureelerfgoed.nl/cho-kennis/rce-art-sample/id/rce_53533
## 9:  https://linkeddata.cultureelerfgoed.nl/cho-kennis/rce-art-sample/id/rce_58790
##                                                description matches
## 1:                                      West-Indisch huis    TRUE
## 2:                                      West-Indisch huis    TRUE
## 3:                                      West-Indisch huis    TRUE
## 4:                                      West-Indisch huis    TRUE
## 5:                                               Batavia    TRUE
## 6:                                               Batavia    TRUE
## 7:                                               Batavia    TRUE
## 8: Emilia van Nassau-Beverweerd (1635-1688), met een bediende    TRUE
## 9:           Wit huis bij maanlicht St. Alban de Montbel    TRUE
##    matched_words
## 1:       Indisch
## 2:       Indisch
```

```
## 3:      Indisch
## 4:      Indisch
## 5:      Batavia
## 6:      Batavia
## 7:      Batavia
## 8:     bediende
## 9:          Wit
```

Colophon

CHAS is created by the Dutch Cultural Heritage Agency, team Thesauri.

The report function provided by CHAS are experimental in nature. It means that it is still under development and may not be free from errors or issues. While we have made every effort to ensure its accuracy, there could be unforeseen challenges and results. Its use is entirely at your discretion and risk.

MIT Licence

Question and remarks: thesauri@cultureelerfgoed.nl