Package ‘childesr’

October 4, 2021

Type Package

Title Accessing the ‘CHILDES’ Database

Description Tools for connecting to ‘CHILDES’, an open repository for transcripts of parent-child interaction. For more information on the underlying data, see <https://childes-db.stanford.edu>.

Version 0.2.2

Depends R (>= 4.0)

License GPL-3

URL https://github.com/langcog/childesr

BugReports https://github.com/langcog/childesr/issues

Encoding UTF-8

Imports DBI (>= 1.1), dbplyr (>= 2.1), dplyr (>= 1.0), jsonlite (>= 1.7), magrittr (>= 2.0), purrr (>= 0.3), RMySQL (>= 0.10.21)

Suggests knitr, rmarkdown, curl

RoxygenNote 7.1.1

VignetteBuilder knitr

NeedsCompilation no

Author Mika Braginsky [aut, cre],
Alessandro Sanchez [aut, ctb],
Daniel Yurovsky [aut],
Kyle MacDonald [ctb],
Stephan Meylan [ctb],
Jessica Mankewitz [ctb]

Maintainer Mika Braginsky <mika.br@gmail.com>

Repository CRAN

Date/Publication 2021-10-04 09:40:02 UTC
connect_to_childes

R topics documented:

- clear_connections ........................................................... 2
- connect_to_childes ......................................................... 2
- get_collections .............................................................. 3
- get_content ................................................................. 4
- get_contexts ................................................................. 5
- get_corpora ................................................................. 6
- get_db_info ................................................................. 7
- get_participants .......................................................... 7
- get_speaker_statistics .................................................... 8
- get_sql_query ............................................................. 9
- get_table ................................................................. 10
- get_tokens ............................................................... 11
- get_transcripts .......................................................... 12
- get_types ................................................................. 13
- get_utterances ........................................................... 14

Index 16

---

### clear_connections

**Clear all MySQL connections**

**Description**

Clear all MySQL connections

**Usage**

```r
clear_connections()
```

---

### connect_to_childes

**Connect to CHILDES**

**Description**

Connect to CHILDES

**Usage**

```r
connect_to_childes(db_version = "current", db_args = NULL)
```

**Arguments**

- **db_version**: String of the name of database version to use
- **db_args**: List with host, user, and password defined
get_collections

Value

con A DBIConnection object for the CHILDES database

Examples

## Not run:
con <- connect_to_childes(db_version = "current", db_args = NULL)
DBI::dbDisconnect(con)

## End(Not run)

get_collections  Get collections

Description

Get collections

Usage

get_collections(connection = NULL, db_version = "current", db_args = NULL)

Arguments

connection A connection to the CHILDES database
db_version String of the name of database version to use
db_args List with host, user, and password defined

Value

A `tbl` of Collection data. If `connection` is supplied, the result remains a remote query, otherwise it is retrieved into a local tibble.

Examples

## Not run:
get_collections()

## End(Not run)
get_content

Description
Get content

Usage

```r
get_content(
  content_type = NULL,
  collection = NULL,
  language = NULL,
  corpus = NULL,
  role = NULL,
  role_exclude = NULL,
  age = NULL,
  sex = NULL,
  target_child = NULL,
  token = NULL,
  stem = NULL,
  part_of_speech = NULL,
  connection
)
```

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>content_type</td>
<td>One of &quot;token&quot; or &quot;utterance&quot;</td>
</tr>
<tr>
<td>collection</td>
<td>A character vector of one or more names of collections</td>
</tr>
<tr>
<td>language</td>
<td>A character vector of one or more languages</td>
</tr>
<tr>
<td>corpus</td>
<td>A character vector of one or more names of corpora</td>
</tr>
<tr>
<td>role</td>
<td>A character vector of one or more roles to include</td>
</tr>
<tr>
<td>role_exclude</td>
<td>A character vector of one or more roles to exclude</td>
</tr>
<tr>
<td>age</td>
<td>A numeric vector of an single age value or a min age value and max age value (inclusive) in months. For a single age value, participants are returned for which that age is within their age range; for two ages, participants are returned for whose age overlaps with the interval between those two ages.</td>
</tr>
<tr>
<td>sex</td>
<td>A character vector of values &quot;male&quot; and/or &quot;female&quot;</td>
</tr>
<tr>
<td>target_child</td>
<td>A character vector of one or more names of children</td>
</tr>
<tr>
<td>token</td>
<td>A character vector of one or more token patterns ('%' matches any number of wildcard characters, '_' matches exactly one wildcard character)</td>
</tr>
<tr>
<td>stem</td>
<td>A character vector of one or more stems</td>
</tr>
<tr>
<td>part_of_speech</td>
<td>A character vector of one or more parts of speech</td>
</tr>
<tr>
<td>connection</td>
<td>A connection to the CHILDES database</td>
</tr>
</tbody>
</table>
get_contexts

Get the utterances surrounding a token(s)

Description
Get the utterances surrounding a token(s)

Usage
get_contexts(
collection = NULL,
language = NULL,
corpus = NULL,
role = NULL,
role_exclude = NULL,
age = NULL,
sex = NULL,
target_child = NULL,
token,
window = c(0, 0),
remove_duplicates = TRUE,
connection = NULL,
db_version = "current",
db_args = NULL)

Arguments

collection A character vector of one or more names of collections
language A character vector of one or more languages
corpus A character vector of one or more names of corpora
role A character vector of one or more roles to include
role_exclude A character vector of one or more roles to exclude
age A numeric vector of an single age value or a min age value and max age value (inclusive) in months. For a single age value, participants are returned for which that age is within their age range; for two ages, participants are returned for whose age overlaps with the interval between those two ages.
sex A character vector of values "male” and/or "female”
target_child A character vector of one or more names of children
token A character vector of one or more token patterns (’%’ matches any number of wildcard characters, ‘.’ matches exactly one wildcard character)
window A length 2 numeric vector of how many utterances before and after each utterance containing the target token to retrieve
get_corpora

Description
Get corpora

Usage
get_corpora(connection = NULL, db_version = "current", db_args = NULL)

Arguments
- connection: A connection to the CHILDES database
- db_version: String of the name of database version to use
- db_args: List with host, user, and password defined

Value
A ‘tbl’ of Corpus data. If ‘connection’ is supplied, the result remains a remote query, otherwise it is retrieved into a local tibble.

Examples
## Not run:
get_corpora()
## End(Not run)
get_db_info

Description
Get information on database connection options

Usage
get_db_info()

Value
List of database info: host name, current version, supported versions, historical versions, username, password

Examples
## Not run:
get_db_info()

## End(Not run)

get_participants

Description
Get participants

Usage
get_participants(
collection = NULL,
corpus = NULL,
target_child = NULL,
role = NULL,
role_exclude = NULL,
age = NULL,
sex = NULL,
connection = NULL,
db_version = "current",
db_args = NULL
)

Arguments

- **collection**: A character vector of one or more names of collections
- **corpus**: A character vector of one or more names of corpora
- **target_child**: A character vector of one or more names of children
- **role**: A character vector of one or more roles to include
- **role_exclude**: A character vector of one or more roles to exclude
- **age**: A numeric vector of a single age value or a min age value and max age value (inclusive) in months. For a single age value, participants are returned for which that age is within their age range; for two ages, participants are returned for whose age overlaps with the interval between those two ages.
- **sex**: A character vector of values "male" and/or "female"
- **connection**: A connection to the CHILDES database
- **db_version**: String of the name of database version to use
- **db_args**: List with host, user, and password defined

Value

A ‘tbl’ of Participant data, filtered down by supplied arguments. If ‘connection’ is supplied, the result remains a remote query, otherwise it is retrieved into a local tibble.

Examples

```r
## Not run:
get_participants()

## End(Not run)
```

---

**get_speaker_statistics**

*Get speaker statistics*

Description

Get speaker statistics

Usage

```r
get_speaker_statistics(
  collection = NULL,
  corpus = NULL,
  target_child = NULL,
  role = NULL,
  role_exclude = NULL,
  age = NULL,
)```
get_sql_query

sex = NULL,
connection = NULL,
db_version = "current",
db_args = NULL
)

Arguments

collection A character vector of one or more names of collections
corpus A character vector of one or more names of corpora
target_child A character vector of one or more names of children
role A character vector of one or more roles to include
role_exclude A character vector of one or more roles to exclude
age A numeric vector of an single age value or a min age value and max age value (inclusive) in months. For a single age value, participants are returned for which that age is within their age range; for two ages, participants are returned for whose age overlaps with the interval between those two ages.
sex A character vector of values "male" and/or "female"
connection A connection to the CHILDES database
db_version String of the name of database version to use
db_args List with host, user, and password defined

Value

A ‘tbl’ of Speaker statistics, filtered down by supplied arguments. If ‘connection’ is supplied, the result remains a remote query, otherwise it is retrieved into a local tibble.

Examples

## Not run:
get_speaker_statistics()

## End(Not run)
Usage

get_sql_query(
    sql_query_string, 
    connection = NULL, 
    db_version = "current", 
    db_args = NULL 
)

Arguments

  sql_query_string  A valid sql query string character
  connection       A connection to the CHILDES database
  db_version       String of the name of database version to use
  db_args          List with host, user, and password defined

Value

  The database after calling the supplied SQL query

Examples

  ## Not run:
  get_sql_query("SELECT * FROM collection")
  
  ## End(Not run)

get_table  Get table

Description

  Get table

Usage

  get_table(connection, name)

Arguments

  connection       A connection to the CHILDES database
  name             String of a table name

Value

  A ‘tbl’
**get_tokens**

Get tokens

**Usage**

```r
get_tokens(
  collection = NULL,
  language = NULL,
  corpus = NULL,
  target_child = NULL,
  role = NULL,
  role_exclude = NULL,
  age = NULL,
  sex = NULL,
  token,
  stem = NULL,
  part_of_speech = NULL,
  replace = TRUE,
  connection = NULL,
  db_version = "current",
  db_args = NULL
)
```

**Arguments**

- `collection`: A character vector of one or more names of collections
- `language`: A character vector of one or more languages
- `corpus`: A character vector of one or more names of corpora
- `target_child`: A character vector of one or more names of children
- `role`: A character vector of one or more roles to include
- `role_exclude`: A character vector of one or more roles to exclude
- `age`: A numeric vector of an single age value or a min age value and max age value (inclusive) in months. For a single age value, participants are returned for which that age is within their age range; for two ages, participants are returned for whose age overlaps with the interval between those two ages.
- `sex`: A character vector of values "male" and/or "female"
- `token`: A character vector of one or more token patterns ("%" matches any number of wildcard characters, "." matches exactly one wildcard character)
- `stem`: A character vector of one or more stems
- `part_of_speech`: A character vector of one or more parts of speech
get_transcripts

```
replace
   A boolean indicating whether to replace "gloss" with "replacement" (i.e. phonologically assimilated form), when available (defaults to TRUE)
connection
   A connection to the CHILDES database
db_version
   String of the name of database version to use
db_args
   List with host, user, and password defined
```

Value

A ‘tbl’ of Token data, filtered down by supplied arguments. If ‘connection’ is supplied, the result remains a remote query, otherwise it is retrieved into a local tibble.

Examples

```r
## Not run:
get_transcripts(token = "dog")
## End(Not run)
```

get_transcripts  Get transcripts

Description

Get transcripts

Usage

```
get_transcripts(
    collection = NULL,
    corpus = NULL,
    target_child = NULL,
    connection = NULL,
    db_version = "current",
    db_args = NULL
)
```

Arguments

```
collection
   A character vector of one or more names of collections
corpus
   A character vector of one or more names of corpora
target_child
   A character vector of one or more names of children
connection
   A connection to the CHILDES database
db_version
   String of the name of database version to use
db_args
   List with host, user, and password defined
```
get_types

Value
A ‘tbl’ of Transcript data, filtered down by supplied arguments. If ‘connection’ is supplied, the result remains a remote query, otherwise it is retrieved into a local tibble.

Examples

```r
## Not run:
get_transcripts()
## End(Not run)
```

---

get_types  Get types

Description
Get types

Usage

```r
get_types(
collection = NULL,
language = NULL,
corpus = NULL,
role = NULL,
role_exclude = NULL,
age = NULL,
sex = NULL,
target_child = NULL,
type = NULL,
connection = NULL,
db_version = "current",
db_args = NULL
)
```

Arguments

- `collection`: A character vector of one or more names of collections
- `language`: A character vector of one or more languages
- `corpus`: A character vector of one or more names of corpora
- `role`: A character vector of one or more roles to include
- `role_exclude`: A character vector of one or more roles to exclude
- `age`: A numeric vector of an single age value or a min age value and max age value (inclusive) in months. For a single age value, participants are returned for which that age is within their age range; for two ages, participants are returned for whose age overlaps with the interval between those two ages.
get_utterances

sex  A character vector of values "male" and/or "female"

target_child  A character vector of one or more names of children

type  A character vector of one or more type patterns (‘’ number of wildcard characters, ‘_’ matches exactly one wildcard character)

connection  A connection to the CHILDES database

db_version  String of the name of database version to use

db_args  List with host, user, and password defined

Value

A ‘tbl’ of Type data, filtered down by supplied arguments. If ‘connection’ is supplied, the result remains a remote query, otherwise it is retrieved into a local tibble.

Examples

```r
## Not run:
get_types()
## End(Not run)
```

---

### get_utterances

**Get utterances**

**Description**

Get utterances

**Usage**

```r
get_utterances(
    collection = NULL,
    language = NULL,
    corpus = NULL,
    role = NULL,
    role_exclude = NULL,
    age = NULL,
    sex = NULL,
    target_child = NULL,
    connection = NULL,
    db_version = "current",
    db_args = NULL
)
```
get_utterances

Arguments

- **collection**: A character vector of one or more names of collections
- **language**: A character vector of one or more languages
- **corpus**: A character vector of one or more names of corpora
- **role**: A character vector of one or more roles to include
- **role_exclude**: A character vector of one or more roles to exclude
- **age**: A numeric vector of an single age value or a min age value and max age value (inclusive) in months. For a single age value, participants are returned for which that age is within their age range; for two ages, participants are returned for whose age overlaps with the interval between those two ages.
- **sex**: A character vector of values "male" and/or "female"
- **target_child**: A character vector of one or more names of children
- **connection**: A connection to the CHILDES database
- **db_version**: String of the name of database version to use
- **db_args**: List with host, user, and password defined

Value

A `tbl` of Utterance data, filtered down by supplied arguments. If `connection` is supplied, the result remains a remote query, otherwise it is retrieved into a local tibble.

Examples

```r
## Not run:
get_utterances(target_child = "Shem")
```

## End(Not run)
Index

clear_connections, 2
connect_to_childes, 2

get_collections, 3
get_content, 4
get_contexts, 5
get_corpora, 6
get_db_info, 7
get_participants, 7
get_speaker_statistics, 8
get_sql_query, 9
get_table, 10
get_tokens, 11
get_transcripts, 12
get_types, 13
get_utterances, 14