

Package ‘CodelistGenerator’

February 14, 2024

Title Identify Relevant Clinical Codes and Evaluate Their Use

Version 2.2.2

Description Generate a candidate code list for the Observational Medical Outcomes Partnership (OMOP) common data model based on string matching. For a given search strategy, a candidate code list will be returned.

License Apache License (>= 2)

Encoding UTF-8

RoxygenNote 7.2.3

Depends R (>= 3.5.0)

Imports CDMConnector (>= 1.3.0), checkmate (>= 2.0.0), DBI (>= 1.1.0), duckdb, dplyr (>= 1.1.0), magrittr (>= 2.0.0), omopgenerics (>= 0.0.2), rlang (>= 1.0.0), glue (>= 1.5.0), stringr (>= 1.4.0), tidyselect (>= 1.2.0), tidyr (>= 1.2.0), cli (>= 3.1.0), purrr, lubridate, PatientProfiles (>= 0.3.0), RJSONIO, vctrs

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

Config/testthat/edition 3

Config/testthat/parallel true

VignetteBuilder knitr

URL <https://darwin-eu.github.io/CodelistGenerator/>

Language en-US

NeedsCompilation no

Author Edward Burn [aut, cre] (<<https://orcid.org/0000-0002-9286-1128>>),
Marti Catala [ctb] (<<https://orcid.org/0000-0003-3308-9905>>)

Maintainer Edward Burn <edward.burn@endorms.ox.ac.uk>

Repository CRAN

Date/Publication 2024-02-14 08:32:36 UTC

R topics documented:

achillesCodeUse	2
codesFromCohort	3
codesFromConceptSet	4
codesInUse	4
compareCodelists	5
findOrphanCodes	6
getATCCodes	8
getCandidateCodes	9
getConceptClassId	10
getDescendants	11
getDomains	12
getDoseForm	12
getDrugIngredientCodes	13
getICD10StandardCodes	14
getMappings	15
getVocabularies	16
getVocabVersion	16
mockVocabRef	17
restrictToCodesInUse	18
sourceCodesInUse	19
summariseCodeUse	19
summariseCohortCodeUse	21

Index	23
--------------	-----------

achillesCodeUse	<i>Summarise code use from achilles counts</i>
-----------------	--

Description

Summarise code use from achilles counts

Usage

```
achillesCodeUse(x, cdm, countBy = c("record", "person"), minCellCount = 5)
```

Arguments

x	Codelist
cdm	cdm_reference via CDMConnector::cdm_from_con()
countBy	Either "record" for record-level counts or "person" for person-level counts
minCellCount	The minimum number of counts to reported, below which results will be suppressed.

Value

A tibble with results

Examples

```
## Not run:
cdm <- mockVocabRef("database")
oa <- getCandidateCodes(cdm = cdm, keywords = "osteoarthritis")
result_achilles <- achillesCodeUse(list(oa = oa$concept_id), cdm = cdm)
result_achilles
CDMConnector::cdmDisconnect(cdm)

## End(Not run)
```

codesFromCohort	<i>Get concept ids from a provided path to cohort json files</i>
-----------------	--

Description

Get concept ids from a provided path to cohort json files

Usage

```
codesFromCohort(path, cdm, withConceptDetails = FALSE)
```

Arguments

path	Path to a file or folder containing JSONs of cohort definitions
cdm	A cdm reference created with CDMConnector
withConceptDetails	If FALSE a vector of concept IDs will be returned for each concept set. If TRUE a tibble will be returned with additional information on the identified concepts.

Value

Named list with concept_ids for each concept set

codesFromConceptSet *Get concept ids from a provided path to json files*

Description

Get concept ids from a provided path to json files

Usage

```
codesFromConceptSet(path, cdm, withConceptDetails = FALSE)
```

Arguments

path Path to a file or folder containing JSONs of concept sets

cdm A cdm reference created with CDMConnector

withConceptDetails If FALSE a vector of concept IDs will be returned for each concept set. If TRUE a tibble will be returned with additional information on the identified concepts.

Value

Named list with concept_ids for each concept set

Examples

```
## Not run:
cdm <- mockVocabRef("database")
x <- codesFromConceptSet(cdm = cdm,
                        path = system.file(package = "CodelistGenerator",
                                           "concepts_for_mock"))
x
CDMConnector::cdmDisconnect(cdm)

## End(Not run)
```

codesInUse *Get codes used in the database*

Description

Get codes used in the database

Usage

```
codesInUse(  
  cdm,  
  table = c("condition_occurrence", "device_exposure", "drug_exposure", "measurement",  
            "observation", "procedure_occurrence", "visit_occurrence")  
)
```

Arguments

cdm	cdm_reference via CDMConnector
table	cdm table

Value

A list of integers indicating codes being used in the database.

Examples

```
## Not run:  
cdm <- mockVocabRef("database")  
x <- codesInUse(cdm = cdm)  
x  
CDMConnector::cdmDisconnect(cdm)  
  
## End(Not run)
```

compareCodelists	<i>Compare two codelists</i>
------------------	------------------------------

Description

Compare two codelists

Usage

```
compareCodelists(codelist1, codelist2)
```

Arguments

codelist1	Output of getCandidateCodes
codelist2	Output of getCandidateCodes

Value

tibble

Examples

```
## Not run:
cdm <- mockVocabRef()
codes1 <- getCandidateCodes(
  cdm = cdm,
  keywords = "Arthritis",
  domains = "Condition",
  includeDescendants = TRUE
)
codes2 <- getCandidateCodes(
  cdm = cdm,
  keywords = c("knee osteoarthritis", "arthrosis"),
  domains = "Condition",
  includeDescendants = TRUE
)
compareCodelists(
  codelist1 = codes1,
  codelist2 = codes2
)
CDMConnector::cdmDisconnect(cdm)

## End(Not run)
```

findOrphanCodes

Find orphan codes related to a codelist

Description

Find orphan codes related to a codelist

Usage

```
findOrphanCodes(
  x,
  cdm,
  domains = "Condition",
  standardConcept = "Standard",
  searchInSynonyms = TRUE,
  searchNonStandard = TRUE,
  includeDescendants = TRUE,
  includeAncestor = TRUE,
  minCellCount = 5
)
```

Arguments

x	Codes for which to find codes related but not included (orphan codes)
cdm	cdm_reference via CDMConnector

domains	Character vector with one or more of the OMOP CDM domain.
standardConcept	Character vector with one or more of "Standard", "Classification", and "Non-standard". These correspond to the flags used for the standard_concept field in the concept table of the cdm.
searchInSynonyms	Either TRUE or FALSE. If TRUE the code will also search using both the primary name in the concept table and synonyms from the concept synonym table.
searchNonStandard	Either TRUE or FALSE. If TRUE the code will also search via non-standard concepts.
includeDescendants	Either TRUE or FALSE. If TRUE descendant concepts of identified concepts will be included in the candidate codelist.
includeAncestor	Either TRUE or FALSE. If TRUE the direct ancestor concepts of identified concepts will be included in the candidate codelist.
minCellCount	The minimum number of counts to reported, below which results will be suppressed. If 0, all results will be reported.

Value

A codelist containing code related to (but not in) the target codelist that are present used in the cdm

Examples

```
## Not run:
cdm <- mockVocabRef("database")
codes <- getCandidateCodes(cdm = cdm,
  keywords = "Musculoskeletal disorder",
  domains = "Condition",
  includeDescendants = FALSE)

orphan_codes <- findOrphanCodes(x = list("msk" = codes$concept_id),
  cdm = cdm,
  domains = "Condition",
  standardConcept = "Standard",
  searchInSynonyms = FALSE,
  searchNonStandard = FALSE,
  includeDescendants = TRUE,
  includeAncestor = FALSE)

orphan_codes
CDMConnector::cdmDisconnect(cdm)

## End(Not run)
```

getATCCodes

Get descendant codes for ATC levels

Description

Get descendant codes for ATC levels

Usage

```
getATCCodes(
  cdm,
  level = c("ATC 1st"),
  name = NULL,
  doseForm = NULL,
  withConceptDetails = FALSE
)
```

Arguments

cdm	cdm_reference via CDMConnector
level	ATC level. Can be one or more of "ATC 1st", "ATC 2nd", "ATC 3rd", "ATC 4th", and "ATC 5th"
name	ATC name of interest. For example, c("Dermatologicals", "Nervous System"), would result in a list of length two with the descendant concepts for these two particular ATC groups.
doseForm	Only descendant codes with the specified dose form will be returned. If NULL, descendant codes will be returned regardless of dose form.
withConceptDetails	If FALSE, each item in the list of results (one per ATC group) will contain a vector of concept IDs for each ingredient. If TRUE each item in the list of results will contain a tibble with additional information on the identified concepts.

Value

A named list, with each item containing a vector of descendant concepts of an ATC group (if withConceptDetails was set as FALSE) or a tibble with the descendant concepts along with additional details about them (if withConceptDetails was set as TRUE).

Examples

```
## Not run:
cdm <- mockVocabRef()
getATCCodes(cdm = cdm, level = "ATC 1st")
CDMConnector::cdmDisconnect(cdm)

## End(Not run)
```

getCandidateCodes	<i>Generate candidate codelist for the OMOP CDM</i>
-------------------	---

Description

This function generates a set of codes that can be considered for creating a phenotype using the OMOP CDM.

Usage

```
getCandidateCodes(
  cdm,
  keywords,
  exclude = NULL,
  domains = "Condition",
  standardConcept = "Standard",
  searchInSynonyms = FALSE,
  searchNonStandard = FALSE,
  includeDescendants = TRUE,
  includeAncestor = FALSE
)
```

Arguments

cdm	cdm_reference via CDMConnector
keywords	Character vector of words to search for. Where more than one word is given (e.g. "knee osteoarthritis"), all combinations of those words should be identified positions (e.g. "osteoarthritis of knee") should be identified.
exclude	Character vector of words to identify concepts to exclude.
domains	Character vector with one or more of the OMOP CDM domain.
standardConcept	Character vector with one or more of "Standard", "Classification", and "Non-standard". These correspond to the flags used for the standard_concept field in the concept table of the cdm.
searchInSynonyms	Either TRUE or FALSE. If TRUE the code will also search using both the primary name in the concept table and synonyms from the concept synonym table.
searchNonStandard	Either TRUE or FALSE. If TRUE the code will also search via non-standard concepts.
includeDescendants	Either TRUE or FALSE. If TRUE descendant concepts of identified concepts will be included in the candidate codelist.
includeAncestor	Either TRUE or FALSE. If TRUE the direct ancestor concepts of identified concepts will be included in the candidate codelist.

Value

tibble

Examples

```
## Not run:
cdm <- CodelistGenerator::mockVocabRef()
CodelistGenerator::getCandidateCodes(
  cdm = cdm,
  keywords = "osteoarthritis"
)
CDMConnector::cdmDisconnect(cdm)

## End(Not run)
```

<code>getConceptClassId</code>	<i>getConceptClassId</i>
--------------------------------	--------------------------

Description

`getConceptClassId`

Usage

```
getConceptClassId(cdm, standardConcept = "Standard", domain = NULL)
```

Arguments

<code>cdm</code>	<code>cdm_reference</code> via <code>CDMConnector</code>
<code>standardConcept</code>	Character vector with one or more of "Standard", "Classification", and "Non-standard". These correspond to the flags used for the <code>standard_concept</code> field in the concept table of the cdm.
<code>domain</code>	Vocabulary domain

Value

The concept class used for a given set of domains

Examples

```
## Not run:
cdm <- mockVocabRef()
getConceptClassId(cdm = cdm, domain = "drug")
CDMConnector::cdmDisconnect(cdm)

## End(Not run)
```

getDescendants	<i>getDescendants</i>
----------------	-----------------------

Description

getDescendants

Usage

```
getDescendants(
  cdm,
  conceptId,
  withAncestor = FALSE,
  ingredientRange = c(0, Inf),
  doseForm = NULL
)
```

Arguments

cdm	cdm_reference via CDMConnector
conceptId	concept_id to search
withAncestor	If TRUE, return column with ancestor. In case of multiple ancestors, concepts will be separated by ";
ingredientRange	Used to restrict descendant codes to those associated with a specific number of drug ingredients. Must be a vector of length two with the first element the minimum number of ingredients allowed and the second the maximum. A value of c(2, 2) would restrict to only concepts associated with two ingredients.
doseForm	Only descendants codes with the specified drug dose form will be returned. If NULL, descendant codes will be returned regardless of dose form.

Value

The descendants of a given concept id

Examples

```
## Not run:
cdm <- mockVocabRef()
getDescendants(cdm = cdm, conceptId = 1)
CDMConnector::cdmDisconnect(cdm)

## End(Not run)
```


Value

The dose forms available for drug concepts

Examples

```
## Not run:
cdm <- mockVocabRef()
getDoseForm(cdm = cdm)
CDMConnector::cdmDisconnect(cdm)

## End(Not run)
```

```
getDrugIngredientCodes
```

Get descendant codes for drug ingredients

Description

Get descendant codes for drug ingredients

Usage

```
getDrugIngredientCodes(
  cdm,
  name = NULL,
  doseForm = NULL,
  ingredientRange = c(1, Inf),
  withConceptDetails = FALSE
)
```

Arguments

cdm	cdm_reference via CDMConnector
name	Names of ingredients of interest. For example, c("acetaminophen", "codeine"), would result in a list of length two with the descendant concepts for these two particular drug ingredients.
doseForm	Only descendants codes with the specified dose form will be returned. If NULL, descendant codes will be returned regardless of dose form.
ingredientRange	Used to restrict descendant codes to those associated with a specific number of ingredients. Must be a vector of length two with the first element the minimum number of ingredients allowed and the second the maximum. A value of c(2, 2) would restrict to only concepts associated with two ingredients.
withConceptDetails	If FALSE, each item in the list of results (one per ingredient) will contain a vector of concept IDs for each ingredient. If TRUE each item in the list of results will contain a tibble with additional information on the identified concepts.

Value

A named list, with each item containing a vector of descendant concepts of an ingredient (if withConceptDetails was set as FALSE) or a tibble with the descendant concepts along with additional details about them (if withConceptDetails was set as TRUE).

Examples

```
## Not run:
cdm <- mockVocabRef()
getDrugIngredientCodes(cdm = cdm, name = "Adalimumab")
CDMConnector::cdmDisconnect(cdm)

## End(Not run)
```

getICD10StandardCodes *Get corresponding standard codes for ICD-10 chapters and sub-chapters*

Description

Get corresponding standard codes for ICD-10 chapters and sub-chapters

Usage

```
getICD10StandardCodes(
  cdm,
  level = c("ICD10 Chapter", "ICD10 SubChapter"),
  name = NULL,
  includeDescendants = TRUE,
  withConceptDetails = FALSE
)
```

Arguments

cdm	cdm_reference via CDMConnector
level	Can be either "ICD10 Chapter" or "ICD10 SubChapter"
name	Name of chapter or sub-chapter of interest. If NULL, all will be considered.
includeDescendants	If FALSE only direct mappings from ICD-10 codes to standard codes will be returned. If TRUE descendants of standard concepts will also be included.
withConceptDetails	If FALSE a vector of concept IDs will be returned for each ICD group If TRUE a tibble will be returned with additional information on the identified concepts.

Value

A named list, with each element containing the corresponding standard codes (and descendants) of ICD chapters and sub-chapters

Examples

```
## Not run:
cdm <- mockVocabRef()
getICD10StandardCodes(cdm = cdm, level = c(
  "ICD10 Chapter",
  "ICD10 SubChapter"
))
CDMConnector::cdmDisconnect(cdm)

## End(Not run)
```

getMappings

Show mappings from non-standard vocabularies to standard

Description

Show mappings from non-standard vocabularies to standard

Usage

```
getMappings(
  candidateCodelist,
  cdm = NULL,
  nonStandardVocabularies = c("ATC", "ICD10CM", "ICD10PCS", "ICD9CM", "ICD9Proc",
    "LOINC", "OPCS4", "Read", "RxNorm", "RxNorm Extension", "SNOMED")
)
```

Arguments

```
candidateCodelist
  Dataframe

cdm
  cdm_reference via CDMConnector::cdm_from_con()

nonStandardVocabularies
  Character vector
```

Value

tibble

Examples

```
## Not run:
cdm <- CodelistGenerator::mockVocabRef()
codes <- CodelistGenerator::getCandidateCodes(
  cdm = cdm,
  keywords = "osteoarthritis"
)
CodelistGenerator::getMappings(
```

```

    cdm = cdm,
    candidateCodelist = codes,
    nonStandardVocabularies = "READ"
  )
  CDMConnector::cdmDisconnect(cdm)

## End(Not run)

```

<code>getVocabularies</code>	<i>getVocabularies</i>
------------------------------	------------------------

Description

`getVocabularies`

Usage

```
getVocabularies(cdm)
```

Arguments

`cdm` `cdm_reference` via `CDMConnector`

Value

Names of available vocabularies

Examples

```

## Not run:
cdm <- mockVocabRef()
getVocabularies(cdm = cdm)
CDMConnector::cdmDisconnect(cdm)

## End(Not run)

```

<code>getVocabVersion</code>	<i>getVocabVersion</i>
------------------------------	------------------------

Description

`getVocabVersion`

Usage

```
getVocabVersion(cdm)
```


Arguments

cdm cdm_reference via CDMConnector

Value

the vocabulary version being used

Examples

```
## Not run:  
cdm <- mockVocabRef()  
getVocabVersion(cdm = cdm)  
CDMConnector::cdmDisconnect(cdm)  
  
## End(Not run)
```

mockVocabRef *Generate example vocabulary database*

Description

Generate example vocabulary database

Usage

```
mockVocabRef(backend = "database")
```

Arguments

backend 'database' (duckdb) or 'data_frame'

Value

cdm reference with mock vocabulary

Examples

```
## Not run:  
cdm <- mockVocabRef()  
cdm  
CDMConnector::cdmDisconnect(cdm)  
  
## End(Not run)
```

restrictToCodesInUse *Filter a codelist to keep only the codes used in the database*

Description

Filter a codelist to keep only the codes used in the database

Usage

```
restrictToCodesInUse(  
  x,  
  cdm,  
  table = c("condition_occurrence", "device_exposure", "drug_exposure", "measurement",  
            "observation", "procedure_occurrence", "visit_occurrence")  
)
```

Arguments

x	A codelist
cdm	cdm_reference via CDMConnector
table	cdm table

Value

A list of integers indicating the codes used in the database

Examples

```
## Not run:  
cdm <- mockVocabRef("database")  
codes <- getCandidateCodes(cdm = cdm,  
                           keywords = "arthritis",  
                           domains = "Condition",  
                           includeDescendants = FALSE)  
x <- restrictToCodesInUse(list("cs1" = codes$concept_id,  
                              "cs2" = 999),  
                          cdm = cdm)  
  
x  
CDMConnector::cdmDisconnect(cdm)  
  
## End(Not run)
```

sourceCodesInUse	<i>Get source codes used in the database</i>
------------------	--

Description

Get source codes used in the database

Usage

```
sourceCodesInUse(  
  cdm,  
  table = c("condition_occurrence", "device_exposure", "drug_exposure", "measurement",  
            "observation", "procedure_occurrence", "visit_occurrence")  
)
```

Arguments

cdm	cdm_reference via CDMConnector
table	cdm table

Value

A list of source codes used in the database.

Examples

```
## Not run:  
cdm <- mockVocabRef("database")  
x <- sourceCodesInUse(cdm = cdm)  
x  
CDMConnector::cdmDisconnect(cdm)  
  
## End(Not run)
```

summariseCodeUse	<i>Summarise code use in patient-level data</i>
------------------	---

Description

Summarise code use in patient-level data

Usage

```
summariseCodeUse(
  x,
  cdm,
  countBy = c("record", "person"),
  byConcept = TRUE,
  byYear = FALSE,
  bySex = FALSE,
  ageGroup = NULL,
  minCellCount = 5
)
```

Arguments

x	List of concept IDs
cdm	cdm_reference via CDMConnector::cdm_from_con()
countBy	Either "record" for record-level counts or "person" for person-level counts
byConcept	TRUE or FALSE. If TRUE code use will be summarised by
byYear	TRUE or FALSE. If TRUE code use will be summarised by year.
bySex	TRUE or FALSE. If TRUE code use will be summarised by sex.
ageGroup	If not NULL, a list of ageGroup vectors of length two.
minCellCount	The minimum number of counts to reported, below which results will be suppressed. If 0, all results will be reported.

Value

A tibble with results overall and, if specified, by strata

Examples

```
## Not run:
con <- DBI::dbConnect(duckdb::duckdb(),
  dbdir = CDMConnector::eunomia_dir())
cdm <- CDMConnector::cdm_from_con(con,
  cdm_schem = "main",
  write_schema = "main")
acetiminophen <- c(1125315, 1127433, 40229134,
  40231925, 40162522, 19133768, 1127078)
poliovirus_vaccine <- c(40213160)
cs <- list(acetiminophen = acetiminophen,
  poliovirus_vaccine = poliovirus_vaccine)
results <- summariseCodeUse(cs, cdm = cdm)
results
CDMConnector::cdmDisconnect(cdm)

## End(Not run)
```

 summariseCohortCodeUse

Summarise code use among a cohort in the cdm reference

Description

Summarise code use among a cohort in the cdm reference

Usage

```
summariseCohortCodeUse(
  x,
  cdm,
  cohortTable,
  cohortId = NULL,
  timing = "any",
  countBy = c("record", "person"),
  byConcept = TRUE,
  byYear = FALSE,
  bySex = FALSE,
  ageGroup = NULL,
  minCellCount = 5
)
```

Arguments

x	Vector of concept IDs
cdm	cdm_reference via CDMConnector::cdm_from_con()
cohortTable	A cohort table from the cdm reference.
cohortId	A vector of cohort IDs to include
timing	When to assess the code use relative cohort dates. This can be "any"(code use any time by individuals in the cohort) or "entry" (code use on individuals' cohort start date).
countBy	Either "record" for record-level counts or "person" for person-level counts
byConcept	TRUE or FALSE. If TRUE code use will be summarised by
byYear	TRUE or FALSE. If TRUE code use will be summarised by year.
bySex	TRUE or FALSE. If TRUE code use will be summarised by sex.
ageGroup	If not NULL, a list of ageGroup vectors of length two.
minCellCount	The minimum number of counts to reported, below which results will be suppressed. If 0, all results will be reported.

Value

A tibble with results overall and, if specified, by strata

Examples

```
## Not run:
con <- DBI::dbConnect(duckdb::duckdb(),
                     dbdir = CDMConnector::eunomia_dir())
cdm <- CDMConnector::cdm_from_con(con,
                                 cdm_schem = "main",
                                 write_schema = "main")
cdm <- CDMConnector::generateConceptCohortSet(cdm = cdm,
conceptSet = list(a = 260139,
                  b = 1127433),
                  name = "cohorts",
                  end = "observation_period_end_date",
                  overwrite = TRUE)

results_cohort_mult <-
summariseCohortCodeUse(list(cs = c(260139,19133873)),
                       cdm = cdm,
                       cohortTable = "cohorts",
                       timing = "entry",
                       minCellCount = 0)

results_cohort_mult
CDMConnector::cdmDisconnect(cdm)

## End(Not run)
```

Index

[achillesCodeUse](#), [2](#)

[codesFromCohort](#), [3](#)

[codesFromConceptSet](#), [4](#)

[codesInUse](#), [4](#)

[compareCodelists](#), [5](#)

[findOrphanCodes](#), [6](#)

[getATCCodes](#), [8](#)

[getCandidateCodes](#), [9](#)

[getConceptClassId](#), [10](#)

[getDescendants](#), [11](#)

[getDomains](#), [12](#)

[getDoseForm](#), [12](#)

[getDrugIngredientCodes](#), [13](#)

[getICD10StandardCodes](#), [14](#)

[getMappings](#), [15](#)

[getVocabularies](#), [16](#)

[getVocabVersion](#), [16](#)

[mockVocabRef](#), [17](#)

[restrictToCodesInUse](#), [18](#)

[sourceCodesInUse](#), [19](#)

[summariseCodeUse](#), [19](#)

[summariseCohortCodeUse](#), [21](#)