Package 'cprd.search'

November 14, 2026

Title Search the Clinical Practice Research Datalink Aurum Dictionary

Version 0.1.2

Description Searches user-supplied Clinical Practice Research Datalink (CPRD) Aurum medical dictionaries to support the development of medical code lists. Provides case-insensitive keyword searches with AND, OR, and exclusion logic, optional regular expressions, approximate matching of complete terms, observation-count filtering, column selection, and deduplicated combinations of results. Exports results to comma-separated files while preserving character identifiers. Includes synthetic example data; real dictionaries must be obtained separately from CPRD <https://www.cprd.com/>.

License MIT + file LICENSE

URL https://github.com/raminrzn/cprd_search

BugReports https://github.com/raminrzn/cprd_search/issues

Encoding UTF-8

RoxygenNote 7.3.2

Depends R (>= 3.5.0)

Imports data.table, stringdist, tibble

Suggests testthat (>= 3.1.5), withr, knitr, rmarkdown

Config/testthat/edition 3

VignetteBuilder knitr

NeedsCompilation no

Author Ramin Rezaeianzadeh [aut, cre]

Maintainer Ramin Rezaeianzadeh <ramin.rezaeianzadeh@outlook.com>

Repository CRAN

Date/Publication 2026-11-14 09:10:20 UTC

Contents

cprd.search-package 2
cprd_browse 3
cprd_combine 3
cprd_export 4
cprd_search 5
cprd_search_exclude 6
cprd_search_fuzzy 7
cprd_search_or 8
load_cprd_dictionary 9

Index **10**

`cprd.search-package` *Search the Clinical Practice Research Datalink Aurum Dictionary*

Description

Load a user-supplied dictionary with `load_cprd_dictionary()`, search with `cprd_search()`, `cprd_search_or()`, `cprd_search_exclude()`, or `cprd_search_fuzzy()`, combine with `cprd_combine()`, and save with `cprd_export()`. `cprd_browse()` provides an interactive console.

Details

Only synthetic example data is included. Real CPRD dictionary files must be obtained separately. Search results are candidates for review, not validated study phenotypes.

Author(s)

Maintainer: Ramin Rezaeianzadeh <ramin.rezaeianzadeh@outlook.com>

See Also

Useful links:

- https://github.com/raminrzn/cprd_search
- Report bugs at https://github.com/raminrzn/cprd_search/issues

cprd_browse *Interactive Dictionary Browser*

Description

Repeatedly search the loaded dictionary at a console prompt. Space-separated keywords use AND logic. Use `OR: asthma, diabetes` for OR logic, or `diabetes -gestational` for exclusions. Prefixes are case-insensitive. Type `q` to quit. Keywords are literal, case-insensitive substrings.

Usage

```
cprd_browse(min_obs = 0)
```

Arguments

`min_obs`	Non-negative whole number giving the minimum observation count.

Value

`NULL`, invisibly, for its printing and prompting side effects.

Examples

```
if (interactive()) {
    load_cprd_dictionary(system.file("extdata", "synthetic-dictionary.tsv",
                                     package = "cprd.search"))
    cprd_browse()
}
```

cprd_combine *Combine Multiple Search Results*

Description

Combine search result tables, keeping the first occurrence of each `MedCodeId`. Columns absent from an input are filled with missing values.

Usage

```
cprd_combine(..., select = NULL)
```

Arguments

`...`	One or more search result tibbles, data.frames, or data.tables. Each must contain a non-missing, non-empty `MedCodeId` column.
`select`	Character vector of columns to keep, or `NULL` for all.

Value

A tibble deduplicated by `MedCodeId`, sorted by decreasing `Observations` when available (missing counts last). Otherwise input order is preserved. Conflicting duplicate rows use the first row, without merging its values with later rows.

Examples

```
load_cprd_dictionary(system.file("extdata", "synthetic-dictionary.tsv",
                                 package = "cprd.search"))
a <- cprd_search("diabetes", select = c("MedCodeId", "Term"))
b <- cprd_search_or("diabetes", "asthma", select = c("MedCodeId", "Term"))
cprd_combine(a, b)
```

---

cprd_export *Export Search Results to CSV*

Description

Save a tibble, data.frame, or data.table of search results to a UTF-8 CSV. Internal `term_lower` and fuzzy `dist` columns are omitted.

Usage

```
cprd_export(results, filename, open = FALSE)
```

Arguments

`results`	A tibble, data.frame, or data.table of search results.
`filename`	Required output path supplied explicitly by the caller. There is no default path. An existing file is overwritten.
`open`	Logical. Open the saved file only when `TRUE` and interactive.

Details

Character identifiers are written without numeric conversion. When importing the CSV, specify character types for identifier columns (for example, `colClasses = "character"` in `utils::read.csv()`) to avoid another application removing leading zeroes or rounding long identifiers.

Value

The supplied file path, invisibly.

Examples

```
load_cprd_dictionary(system.file("extdata", "synthetic-dictionary.tsv",
                                 package = "cprd.search"))
results <- cprd_search("diabetes")
file <- tempfile(fileext = ".csv")
cprd_export(results, file)
utils::read.csv(file, colClasses = "character")
unlink(file)
```

---

cprd_search *Search the CPRD Dictionary with AND Logic*

Description

Find rows containing every supplied keyword. Matching is case-insensitive and uses literal substrings unless `regex = TRUE`.

Usage

```
cprd_search(..., min_obs = 0, column = "Term", select = NULL, regex = FALSE)
```

Arguments

`...`	One or more character strings or character vectors of keywords. Empty strings and missing values are not accepted.
`min_obs`	Non-negative whole number giving the minimum observation count. Zero includes rows with missing counts; positive values exclude them.
`column`	Name of the dictionary column to search. Default is `"Term"`.
`select`	Character vector of column names to return, or `NULL` for all dictionary columns except the internal `term_lower` column.
`regex`	Logical. Interpret keywords as regular expressions if `TRUE`. The default is `FALSE`, so punctuation in medical codes is literal.

Value

A tibble, sorted by decreasing `Observations`, with missing counts last. No matches returns a zero-row tibble with the same columns and types.

Examples

```
load_cprd_dictionary(system.file("extdata", "synthetic-dictionary.tsv",
                                 package = "cprd.search"))
cprd_search("family history", "lung cancer")
cprd_search("diabetes", min_obs = 50, select = c("MedCodeId", "Term"))
cprd_search("H33..", column = "CleansedReadCode")
cprd_search("^asthma$", regex = TRUE)
```

cprd_search_exclude *Search with Exclusions*

Description

Find rows matching inclusion keywords but none of the exclusion keywords. Matching is case-insensitive and literal unless `regex = TRUE`.

Usage

```
cprd_search_exclude(
  include,
  exclude = NULL,
  min_obs = 0,
  or_logic = FALSE,
  select = NULL,
  column = "Term",
  regex = FALSE
)
```

Arguments

include	Non-empty character vector of inclusion keywords.
exclude	Character vector of exclusion keywords, or `NULL`.
min_obs	Non-negative whole number giving the minimum observation count. Zero includes rows with missing counts; positive values exclude them.
or_logic	Logical. Match any inclusion keyword if `TRUE`, or every inclusion keyword if `FALSE` (the default).
select	Character vector of column names to return, or `NULL` for all dictionary columns except the internal `term_lower` column.
column	Name of the dictionary column to search. Default is `"Term"`.
regex	Logical. Interpret keywords as regular expressions if `TRUE`. The default is `FALSE`, so punctuation in medical codes is literal.

Value

A tibble, sorted by decreasing `Observations`, with missing counts last. No matches returns a zero-row tibble with the same columns and types.

Examples

```
load_cprd_dictionary(system.file("extdata", "synthetic-dictionary.tsv",
                                 package = "cprd.search"))
cprd_search_exclude("diabetes", exclude = "gestational")
```

`cprd_search_fuzzy` *Fuzzy Search Complete Dictionary Terms*

Description

Compare a query with complete, case-insensitive term descriptions using string distances. Every eligible term is compared; no prefix filter is used.

Usage

```
cprd_search_fuzzy(
  term,
  max_dist = NULL,
  method = "osa",
  min_obs = 0,
  max_results = 50,
  select = NULL
)
```

Arguments

`term`	One non-empty character string to compare with `Term`.
`max_dist`	Maximum distance to accept. `NULL` uses 2 for edit/count distances and 0.1 for `"jw"`, `"cosine"`, or `"jaccard"`. For these normalized methods the threshold must be between 0 and 1.
`method`	Distance method: `"osa"` (default), `"lv"`, `"dl"`, `"hamming"`, `"lcs"`, `"qgram"`, `"cosine"`, `"jaccard"`, or `"jw"`. See stringdist::stringdist-metrics for definitions. `"jw"` uses Jaro distance with the default prefix weight of zero.
`min_obs`	Non-negative whole number giving the minimum observation count. Zero includes rows with missing counts; positive values exclude them.
`max_results`	Non-negative whole number limiting the result count. Zero returns a correctly typed, empty result.
`select`	Character vector of column names to return, or `NULL` for all dictionary columns except the internal `term_lower` column.

Details

Distances compare the whole query with the whole description, not individual words or substrings. Thus `"diabeties"` can match `"diabetes"` at an edit distance of 1, but a longer description such as `"gestational diabetes"` requires a larger edit threshold. Use `cprd_search()` for literal substring searches. Search candidates still require review when building a study code list.

Value

A tibble containing dictionary columns and numeric `dist`, ordered by increasing distance, then decreasing observations. `select` can include `dist`. Missing or empty descriptions do not match.

Examples

```
load_cprd_dictionary(system.file("extdata", "synthetic-dictionary.tsv",
                                 package = "cprd.search"))
cprd_search_fuzzy("diabeties", select = c("MedCodeId", "Term", "dist"))
cprd_search_fuzzy("xiabetes", max_dist = 1)
cprd_search_fuzzy("diabeties", method = "jw", max_dist = 0.1)
```

`cprd_search_or` *Search the CPRD Dictionary with OR Logic*

Description

Find rows containing any supplied keyword, using case-insensitive matching.

Usage

```
cprd_search_or(..., min_obs = 0, column = "Term", select = NULL, regex = FALSE)
```

Arguments

`...`	One or more character strings or character vectors of keywords. Empty strings and missing values are not accepted.
`min_obs`	Non-negative whole number giving the minimum observation count. Zero includes rows with missing counts; positive values exclude them.
`column`	Name of the dictionary column to search. Default is `"Term"`.
`select`	Character vector of column names to return, or `NULL` for all dictionary columns except the internal `term_lower` column.
`regex`	Logical. Interpret keywords as regular expressions if `TRUE`. The default is `FALSE`, so punctuation in medical codes is literal.

Value

A tibble, sorted by decreasing `Observations`, with missing counts last. No matches returns a zero-row tibble with the same columns and types.

Examples

```
load_cprd_dictionary(system.file("extdata", "synthetic-dictionary.tsv",
                                 package = "cprd.search"))
cprd_search_or("asthma", "lung cancer", select = c("MedCodeId", "Term"))
```

load_cprd_dictionary *Load the CPRD Aurum Medical Dictionary*

Description

Read a user-supplied tab-delimited dictionary and cache it for searching. The package includes only a small, entirely synthetic example dictionary.

Usage

```
load_cprd_dictionary(filepath = NULL, force = FALSE)
```

Arguments

filepath	Path to a dictionary file. With NULL, reuse the cached dictionary or look for CPRDAurumMedical.txt in the working directory.
force	Logical. Reload the file even when its path is already cached.

Details

Required columns are MedCodeId, Term, and Observations. Standard additional columns are OriginalReadCode, CleansedReadCode, SnomedCTConceptId, SnomedCTDescriptionId, Release, and EmisCodeCategoryId. All columns except Observations are read as character to preserve identifiers. Observation counts must be non-negative whole numbers; blank or NA counts are treated as missing. Missing terms do not match searches. Empty or missing medical code identifiers are rejected.

Supplying a different path loads a new dictionary. Failed loads leave the previous cache intact. Use force = TRUE after changing a cached file. Real dictionaries must be obtained separately through the Clinical Practice Research Datalink; see https://www.cprd.com/ for access information.

Value

An independent copy of the loaded data.table, invisibly. Identifier columns are character, Observations is numeric, and term_lower is an internal lowercase copy of Term.

Examples

```
path <- system.file("extdata", "synthetic-dictionary.tsv", package = "cprd.search")
dict <- load_cprd_dictionary(path)
nrow(dict)
```

Index

cprd.search *(cprd.search-package)*, 2
cprd.search-package, 2
cprd_browse, 3
cprd_browse(), *2*
cprd_combine, 3
cprd_combine(), *2*
cprd_export, 4
cprd_export(), *2*
cprd_search, 5
cprd_search(), *2*, *7*
cprd_search_exclude, 6
cprd_search_exclude(), *2*
cprd_search_fuzzy, 7
cprd_search_fuzzy(), *2*
cprd_search_or, 8
cprd_search_or(), *2*

load_cprd_dictionary, 9
load_cprd_dictionary(), *2*

stringdist::stringdist-metrics, *7*

utils::read.csv(), *4*