

Package ‘ConceptSetConstructionEvaluation’

September 10, 2026

Type Package
Title Evaluating Concept Set Constructors
Version 0.1.0
Date 2026-09-11
Description Evaluating concept set constructors using various gold standards.
Depends R (>= 4.0.0)
Imports dplyr,
openxlsx,
checkmate,
rlang
Suggests testthat,
License Apache License
URL <https://ohdsi.github.io/DatabaseConnector/>, <https://github.com/OHDSI/DatabaseConnector>
BugReports <https://github.com/OHDSI/DatabaseConnector/issues>
Roxygen list(markdown = TRUE)
Encoding UTF-8
Config/roxygen2/version 8.1.0

Contents

ai3ConceptSets	2
evaluateConceptAdjudication	2
evaluateConceptSets	3
getConceptSetTargets	3
getConceptsForAdjudication	4
Index	5

ai3ConceptSets

AI3 Concept Sets

Description

Concept sets as created by AI3 in the MindsMeetMachines challenge. The dataset includes the set of concept IDs included in each of the six concept sets according to AI3.

Usage

```
data(ai3ConceptSets)
```

Format

a dataframe with 681 rows and 2 variables

- id : The ID of the concept set
- conceptId: The concept ID of a concept included in the concept set

evaluateConceptAdjudication

Evaluate concept adjudication

Description

The gold standard consists of 52 concept set targets. For each concept set target, A combination of vocabulary lookup and OHDSI's Pheobe 2.0 tool were used to find possibly relevant concepts, of which 25 were randomly sampled. In addition, LLMs were used to generate clinical definitions of each concept set target.

The 1,300 target-definition-concept triplets were manually reviewed to determine whether the concept should be part of the concept set according to the definition.

PPV, sensitivity, and specificity are computed overall, and stratified by concept set target domain.

Usage

```
evaluateConceptAdjudication(concepts)
```

Arguments

concepts	The data frame returned by getConceptsForAdjudication() with an extra 'adjudication' column, having value 'YES' if the concept should be included in a concept set for the target, or 'NO' otherwise.
----------	---

Value

A data frame with performance statistics.

Examples

```
concepts <- getConceptsForAdjudication()

# Using random adjudication for this example:
concepts$adjudication <- sample(c("YES", "NO"), nrow(concepts), replace = TRUE)

evaluateConceptAdjudication(concepts)
```

evaluateConceptSets	<i>Evaluate concept sets</i>
---------------------	------------------------------

Description

Evaluate concept sets

Usage

```
evaluateConceptSets(conceptSets)
```

Arguments

conceptSets	The concept sets to evaluate. A data frame with two columns: id: the ID of the concept set, and conceptId: the concept ID of the concept included in the concept set. Can contain 1 or more rows per concept set.
-------------	---

Value

The evaluation of the concept sets

Examples

```
data(ai3ConceptSets)
results <- evaluateConceptSets(ai3ConceptSets)
results$f1ConservativeWeighted[results$id == "Total"]
```

getConceptSetTargets	<i>Get the concept set targets included in the concept sets</i>
----------------------	---

Description

Get the concept set targets included in the concept sets

Usage

```
getConceptSetTargets()
```

Value

A data frame with one row per concept set, and the following columns:

- **id**: A unique identifier. Use this when feeding the results in `evaluateConceptSets()`.
- **name**: The name of the clinical idea for which to generate the concept set.
- **definition**: A long free-text definition of the clinical idea, including the intended use of the concept set.
- **closestConceptId**: The concept ID of the concept that most closely matched the clinical idea.
- **closestConceptName**: The name of the concept that most closely matched the clinical idea.

Examples

```
getConceptSetTargets()
```

```
getConceptsForAdjudication
```

Get concepts for concept adjudication evaluation

Description

Get a set of concept set targets and candidate concepts for evaluating the process of concept adjudication (determining) whether a concept belongs to a concept set.

Use the `evaluateConceptAdjudication()` function to evaluate your adjudications of these concepts.

Usage

```
getConceptsForAdjudication()
```

Value

A data frame with the following columns:

- **targetName**: The name of the target for the concept set (what the concept set should be about).
- **targetDefinition**: A detailed description of the concept set target.
- **conceptId**: The concept ID of the candidate concept.
- **conceptName**: The name of the candidate concept.
- **vocabularyId**: The vocabulary ID of the candidate concept.
- **domainId**: The domain ID of the candidate concept.
- **conceptClassId**: The concept class ID of the candidate concept.

Index

* **datasets**

ai3ConceptSets, [2](#)

ai3ConceptSets, [2](#)

evaluateConceptAdjudication, [2](#)

evaluateConceptSets, [3](#)

getConceptSetTargets, [3](#)

getConceptsForAdjudication, [4](#)