

Package: cora (via r-universe)

September 12, 2024

Title Cora Data for Entity Resolution

Version 0.1.0

Description Duplicated publication data (pre-processed and formatted) for entity resolution. This data set contains a total of 1879 records. The following variables are included in the data set: id, title, book title, authors, address, date, year, editor, journal, volume, pages, publisher, institution, type, tech, note. The data set has a respective gold data set that provides information on which records match based on id.

URL <https://github.com/resteorts/cora>

BugReports <https://github.com/resteorts/cora/issues>

Depends R (>= 3.4.0)

License CC0

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1.9000

Repository <https://cleanzr.r-universe.dev>

RemoteUrl <https://github.com/cleanzr/cora>

RemoteRef HEAD

RemoteSha 70e32d5d2e947407015de334a99bfaf35138d3ed

Contents

cora	2
cora_gold	2
cora_gold_update	3

Index	4
--------------	----------

cora	<i>CORA data set</i>
------	----------------------

Description

This provides a record linkage data set with information about different CORA research papers.

Usage

```
cora
```

Format

A data frame with 16 variables: id, title, book_title, authors, address, date, year, editor, journal, volume, pages, publisher, institution, type, tech, note.

This data set includes 1879 CORA research papers. It is appropriate for performing various types of record linkage and can be assessed by standard record linkage methods.

Examples

```
head(cora)
dim(cora)
```

cora_gold	<i>Cora Gold</i>
-----------	------------------

Description

This data set includes the matched record pairs based on ID.

Usage

```
cora_gold
```

Format

A data frame with 2 variables: id1, id2

This data set includes the matched record pairs based on ID from the CORA data set. This data set can be used to evaluate the performance of record linkage methods performed on the CORA data set.

Examples

```
head(cora_gold)
dim(cora_gold)
```

cora_gold_update	<i>Cora Gold Update</i>
------------------	-------------------------

Description

This data set includes the matched record pairs based on ID.

Usage

cora_gold_update

Format

A data frame with 2 variables: cora_id, unique_id

This data set includes the matched record pairs based on ID from the CORA data set. This data set can be used to evaluate the performance of record linkage methods performed on the CORA data set.

Examples

```
head(cora_gold_update)
dim(cora_gold_update)
```

Index

* datasets

cora, [2](#)

cora_gold, [2](#)

cora_gold_update, [3](#)

cora, [2](#)

cora_gold, [2](#)

cora_gold_update, [3](#)