Package 'cd'

October 12, 2022

Title CD Data for Entity Resolution

Version 0.1.0

Description Duplicated music data (pre-processed and formatted) for entity resolution. The total size of the data set is 9763. There are respective gold standard records that are labeled and can be considered as a unique identifier.

URL

BugReports

Depends R (>= 3.4.0)

License CC0

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1.9000

NeedsCompilation no

Author Rebecca Steorts [aut, cre],
Andee Kaplan [aut],
Srini. Sunil [aut]

Maintainer Rebecca Steorts <beka@stat.duke.edu>

Repository CRAN

Date/Publication 2020-10-22 08:40:02 UTC

R topics documented:

---

cd   *CD Data Set*

Description

This data set includes 9763 CDs randomly extracted from freeDB.

Usage

```
cd
```

Format

A data frame with 11 variables: `pk`, `id`, `artist`, `title`, `category`, `genre`, `cdextra`, `year`, `track_number`, `song_name`

This data set includes 9763 CDs randomly extracted from freeDB. It is appropriate for performing various types of record linkage and can be assessed by standard record linkage methods.

Examples

```
head(cd)
dim(cd)
```

---

cd_gold   *CD Gold*

Description

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

Usage

```
cd_gold
```

Format

A data frame with 2 variables: `disc1_id`, `disc2_id`

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

Examples

```
head(cd_gold)
dim(cd_gold)
```

Index