

Package ‘LinkIt’

June 12, 2019

Title An Algorithm for Dataset Linkage

Version 0.0

Authors 'Brian Libgober <libgober@gmail.com> [aut,cre], Connor Jerzak <cjerzak@g.harvard.edu> [aut, cre]'

Description An R package for estimating record linkage.

Depends R (>= 3.3.3)

License

Creative Commons Attribution-Noncommercial-No Derivative Works 4.0, for academic use only.

Encoding UTF-8

LazyData no

Maintainer 'Connor Jerzak' <connor.jerzak@gmail.com>

Imports FNN

RoxygenNote 6.1.0

R topics documented:

LinkIt 1

Index 3

| | |
|--------|---------------|
| LinkIt | <i>LinkIt</i> |
|--------|---------------|

Description

Record linkage description.

Usage

LinkIt(x,y,by,...)

Arguments

| | |
|----------------|--|
| x, y | data frames to be merged |
| by, by.x, by.y | specifications of the columns used for merging. |
| fuzzy_step | A Boolean (T/F) indicating whether an additional pre-processing step with the LinkedThem database. Default option is TRUE. |
| control | A list specifying how to process the alias text. See “Details”. |

Details

LinkIt can automatically process the alias text for each dataset. In 'Control', users can specify the following options.

Set 'RemoveCommonWords' to TRUE to remove common words (those appearing in > 10% of aliases).

Set 'NormalizeSpaces' to TRUE to remove hanging whitespaces.

Set 'RemovePunctuation' to TRUE to remove punctuation.

Set 'ToLower' to TRUE to ignore case.

Set 'PreprocessingFuzzyThreshold' to some number between 0 and 1 to specify the threshold for the pre-processing fuzzy matching step.

Value

z The merged data frame.

Index

LinkIt, [1](#)