# Package 'LinkIt'

March 15, 2021

Encoding UTF-8  LazyData true  Maintainer 'Connor Jerzak' <connor.jerzak@gmail.com> Imports data.table,plyr  RoxygenNote 7.1.1  R topics documented:  FastFuzzyMatch LinkIt  Index  FastFuzzyMatch FastFuzzyMatch  PastFuzzyMatch  FastFuzzyMatch  Usage</connor.jerzak@gmail.com>	Water 13, 2021
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 (&gt;= 3.3.3)  License  Creative Commons Attribution-Noncommercial-No Derivative Works 4.0, for academic use on Encoding UTF-8  LazyData true  Maintainer 'Connor Jerzak' <connor.jerzak@gmail.com> Imports data.table.plyr  RoxygenNote 7.1.1  R topics documented:  FastFuzzyMatch  Linklt  Index  FastFuzzyMatch  Pescription  Record linkage description.  Usage</connor.jerzak@gmail.com></cjerzak>	Title An Algorithm for Dataset Linkage
nor Jerzak <cjerzak@g.harvard.edu> [aut, cre]'  Description An R package for estimating record linkage.  Depends R (&gt;= 3.3.3)  License  Creative Commons Attribution-Noncommercial-No Derivative Works 4.0, for academic use on Encoding UTF-8  LazyData true  Maintainer 'Connor Jerzak' <connor.jerzak@gmail.com>  Imports data.table.plyr  RoxygenNote 7.1.1  R topics documented:  FastFuzzyMatch  LinkIt  Index  FastFuzzyMatch  FastFuzzyMatch  Description  Record linkage description.  Usage</connor.jerzak@gmail.com></cjerzak@g.harvard.edu>	Version 0.0
Depends R (>= 3.3.3)  License Creative Commons Attribution-Noncommercial-No Derivative Works 4.0, for academic use on Encoding UTF-8  LazyData true Maintainer 'Connor Jerzak' <connor.jerzak@gmail.com> Imports data.table,plyr RoxygenNote 7.1.1  R topics documented: FastFuzzyMatch LinkIt Index  FastFuzzyMatch  FastFuzzyMatch  Record linkage description.  Usage</connor.jerzak@gmail.com>	
License Creative Commons Attribution-Noncommercial-No Derivative Works 4.0, for academic use on Encoding UTF-8  LazyData true  Maintainer 'Connor Jerzak' <connor.jerzak@gmail.com> Imports data.table,plyr  RoxygenNote 7.1.1  R topics documented: FastFuzzyMatch LinkIt  Index  FastFuzzyMatch  FastFuzzyMatch  Description Record linkage description.  Usage</connor.jerzak@gmail.com>	<b>Description</b> An R package for estimating record linkage.
Creative Commons Attribution-Noncommercial-No Derivative Works 4.0, for academic use on Encoding UTF-8  LazyData true  Maintainer 'Connor Jerzak' <connor.jerzak@gmail.com> Imports data.table,plyr  RoxygenNote 7.1.1  R topics documented:  FastFuzzyMatch LinkIt  Index  FastFuzzyMatch  Description  Record linkage description.  Usage</connor.jerzak@gmail.com>	<b>Depends</b> R (>= $3.3.3$ )
LazyData true  Maintainer 'Connor Jerzak' <connor.jerzak@gmail.com> Imports data.table,plyr  RoxygenNote 7.1.1  R topics documented:  FastFuzzyMatch LinkIt  Index  FastFuzzyMatch FastFuzzyMatch  Description  Record linkage description.  Usage</connor.jerzak@gmail.com>	License Creative Commons Attribution-Noncommercial-No Derivative Works 4.0, for academic use only
Maintainer 'Connor Jerzak' <connor.jerzak@gmail.com> Imports data.table,plyr RoxygenNote 7.1.1  R topics documented:  FastFuzzyMatch LinkIt  Index  FastFuzzyMatch FastFuzzyMatch  PastFuzzyMatch  FastFuzzyMatch  Usage</connor.jerzak@gmail.com>	Encoding UTF-8
Imports data.table,plyr  RoxygenNote 7.1.1  R topics documented:  FastFuzzyMatch	LazyData true
RoxygenNote 7.1.1  R topics documented:  FastFuzzyMatch LinkIt  Index  FastFuzzyMatch  FastFuzzyMatch  Description  Record linkage description.  Usage	Maintainer 'Connor Jerzak' <connor.jerzak@gmail.com></connor.jerzak@gmail.com>
R topics documented:  FastFuzzyMatch LinkIt  Index  FastFuzzyMatch  FastFuzzyMatch  Description Record linkage description.  Usage	Imports data.table,plyr
FastFuzzyMatch LinkIt  Index  FastFuzzyMatch  FastFuzzyMatch  Description  Record linkage description.  Usage	RoxygenNote 7.1.1
FastFuzzyMatch  Description  Record linkage description.  Usage	FastFuzzyMatch
Description  Record linkage description.  Usage	Index
Record linkage description.  Usage	FastFuzzyMatch FastFuzzyMatch
Usage	
	Record linkage description.
Foot Furzy Motob (v. v. by	Usage
rastruzzymatch(x,y,by,)	FastFuzzyMatch(x,y,by,)

2 LinkIt

#### **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.

LinkIt LinkIt

## Description

Record linkage description.

### Usage

## **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".

LinkIt 3

### **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.

### Value

z The merged data frame.

# Index

FastFuzzyMatch, 1

LinkIt, 2