

# Package ‘LinkIt’

March 14, 2021

**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** true

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

**Imports** data.table,plyr

**RoxygenNote** 7.1.1

## R topics documented:

FastFuzzyMatch . . . . .	1
LinkIt . . . . .	2

<b>Index</b>	<b>4</b>
--------------	----------

---

FastFuzzyMatch	<i>FastFuzzyMatch</i>
----------------	-----------------------

---

## Description

Record linkage description.

## Usage

FastFuzzyMatch(x,y,by,...)

**Arguments**

<code>x, y</code>	data frames to be merged
<code>by, by.x, by.y</code>	specifications of the columns used for merging.
<code>fuzzy_step</code>	A Boolean (T/F) indicating whether an additional pre-processing step with the LinkedThem database. Default option is TRUE.
<code>control</code>	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	<i>LinkIt</i>
--------	---------------

---

**Description**

Record linkage description.

**Usage**

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

**Arguments**

<code>x, y</code>	data frames to be merged
<code>by, by.x, by.y</code>	specifications of the columns used for merging.
<code>fuzzy_step</code>	A Boolean (T/F) indicating whether an additional pre-processing step with the LinkedThem database. Default option is TRUE.
<code>control</code>	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.

**Value**

z The merged data frame.

# Index

FastFuzzyMatch, [1](#)

LinkIt, [2](#)