

# Package ‘diffrprojects’

September 2, 2016

**Title** Using diffr for more than two files

**Date** 2016-08-26

**Version** 0.1.4.90000

**Description** This is a description still to be done but to  
prevent checks about complaining about too short descriptions  
this does not simply read TBD.

**Depends** R (>= 3.0.0), stringb (>= 0.1.10), rtext (>= 0.1.10)

**License** MIT + file LICENSE

**LazyData** TRUE

**Imports** R6 (>= 2.1.2), hellno (>= 0.0.1), dplyr(>= 0.5.0), data.table  
(>= 1.9.6), dtplyr (>= 0.0.1), Rcpp (>= 0.12.6), stringdist (>= 0.9.4.1), magrittr, stats, graphics

**Suggests** testthat, knitr, rmarkdown

**BugReports** <https://github.com/petermeissner/diffrprojects/issues>

**URL** <https://github.com/petermeissner/diffrprojects>

**RoxygenNote** 5.0.1

**VignetteBuilder** knitr

**LinkingTo** Rcpp

**Author** Peter Meissner [aut, cre],  
Ulrich Sieberer [cph],  
University of Konstanz [cph]

**Maintainer** Peter Meissner <retep.meissner@gmail.com>

**Archs** x64

**R topics documented:**

---

```
as.data.frame.alignment_list
```

*as.data.frame method for for named lists of data.frames*

---

### Description

as.data.frame method for for named lists of data.frames

### Usage

```
## S3 method for class 'alignment_list'
as.data.frame(x, row.names = NULL,
  optional = FALSE, ...)
```

### Arguments

x	any R object.
row.names	NULL or a character vector giving the row names for the data frame. Missing values are not allowed.
optional	logical. If TRUE, setting row names and converting column names (to syntactic names: see <a href="#">make.names</a> ) is optional. Note that all of R's <b>base</b> package as.data.frame() methods use optional only for column names treatment, basically with the meaning of <a href="#">data.frame</a> (*, check.names = !optional).
...	additional arguments to be passed to or from methods.

---

```
as.data.frame.named_df_list
```

*as.data.frame method for for named lists of data.frames*

---

### Description

as.data.frame method for for named lists of data.frames

### Usage

```
## S3 method for class 'named_df_list'
as.data.frame(x, row.names = NULL, optional = FALSE,
  dfnamevar = "name", ...)
```

### Arguments

x	any R object.
row.names	NULL or a character vector giving the row names for the data frame. Missing values are not allowed.
optional	logical. If TRUE, setting row names and converting column names (to syntactic names: see <a href="#">make.names</a> ) is optional. Note that all of R's <b>base</b> package as.data.frame() methods use optional only for column names treatment, basically with the meaning of <a href="#">data.frame</a> (*, check.names = !optional).
dfnamevar	in which variable should list item names be saved
...	additional arguments to be passed to or from methods.

---

choose_options	<i>(choose from a number of pre-sorted options) takes a vector pair of toki1 / toki2 and a vector pair of res_token_i_1 / res_token_i_2 and chooses so that each 1st and exh 2nd value only is used where res_token_i_x identiefies already used items.</i>
----------------	---

---

### Description

(choose from a number of pre-sorted options) takes a vector pair of toki1 / toki2 and a vector pair of res\_token\_i\_1 / res\_token\_i\_2 and chooses so that each 1st and exh 2nd value only is used where res\_token\_i\_x identiefies already used items.

### Usage

```
choose_options(toki1, toki2, res_token_i_1, res_token_i_2)
```

### Arguments

toki1	first number of number pair to choose from
toki2	second number of number pair to choose from
res_token_i_1	already used first numbers
res_token_i_2	already used second numbers // @keywords internal

---

diffrrproject	<i>class for diffrrproject</i>
---------------	--------------------------------

---

### Description

class for diffrrproject  
class for dp\_align  
class for dp\_base

### Usage

```
diffrrproject  
  
dp_align  
  
dp_base
```

### Format

[R6Class](#) object.

### Value

Object of [diffrrproject](#)  
Object of [dp\\_align](#)  
Object of [dp\\_base](#)

diff\_align

*aligning texts***Description**

Function aligns two texts side by side as a data.frame with change type and distance given as well

**Usage**

```
diff_align(text1 = NULL, text2 = NULL, tokenizer = NULL, ignore = NULL,
  clean = NULL, distance = c("lv", "osa", "dl", "hamming", "lcs", "qgram",
    "cosine", "jaccard", "jw", "soundex"), useBytes = FALSE, weight = c(d = 1,
    i = 1, s = 1, t = 1), maxDist = 0, q = 1, p = 0,
  nthread = getOption("sd_num_thread"), verbose = TRUE, ...)
```

**Arguments**

text1	first text
text2	second text
tokenizer	defaults to NULL which will trigger linewise tokenization; accepts a function that turns a text into a token data frame; a token data frame has at least three columns: from (first character of token), to (last character of token) token (the token)
ignore	defaults to NULL which means that nothing is ignored; function that accepts a token data frame (see above) and returns a possibly subseted data frame of hte same form
clean	defaults to NULL which means that nothing cleaned; accepts a function that takes a vector of tokens and returns a vector of same length - potentially clean up
distance	defaults to Levenshtein ("lv"); see <a href="#">amatch</a> , <a href="#">stringdist-metrics</a> , <a href="#">stringdist</a>
useBytes	Perform byte-wise comparison, see <a href="#">stringdist-encoding</a> .
weight	For method='osa' or 'dl', the penalty for deletion, insertion, substitution and transposition, in that order. When method='lv', the penalty for transposition is ignored. When method='jw', the weights associated with characters of a, characters from b and the transposition weight, in that order. Weights must be positive and not exceed 1. weight is ignored completely when method='hamming', 'qgram', 'cosine', 'Jaccard', 'lcs', or soundex.
maxDist	[DEPRECATED AND WILL BE REMOVED 2016] Currently kept for backward compatibility. It does not offer any speed gain. (In fact, it currently slows things down when set to anything different from Inf).
q	Size of the $q$ -gram; must be nonnegative. Only applies to method='qgram', 'jaccard' or 'cosine'.
p	Penalty factor for Jaro-Winkler distance. The valid range for p is $0 \leq p \leq 0.25$ . If p=0 (default), the Jaro-distance is returned. Applies only to method='jw'.
nthread	Maximum number of threads to use. By default, a sensible number of threads is chosen, see <a href="#">stringdist-parallelization</a> .
verbose	should function report on its doings via messages or not
...	further arguments passed through to distance function

**Value**

dataframe with tokens aligned according to distance

---

dp_text_base_data	<i>function providing basic information on texts within diffproject</i>
-------------------	---

---

**Description**

function providing basic information on texts within diffproject

**Usage**

```
dp_text_base_data(dp)
```

**Arguments**

dp	a diffproject object
----	----------------------

---

dummyimport	<i>imports</i>
-------------	----------------

---

**Description**

imports

**Usage**

```
dummyimport()
```