Package ‘edm1’

June 20, 2024

Title Set of functions to work with pairs in character

Version 2.0.0.0

Description Provides functions to detect the pairs of elements in a character, to merge the indexes of two type of pairs from the same character, to give pairs to a character according to a special algorytm...

License GPL (==3)

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.1

Imports stringr,
stringi,
dplyr,
openxlsx

Contents

`depth_pairs_findr` *depth_pairs_findr*

Description

Takes the pair vector as an input and associate to each pair a level of depth, see examples

Usage

```
depth_pairs_findr(inpt)
```

Arguments

`inpt`	is the pair vector

Examples

```
print(depth_pairs_findr(c(1, 1, 2, 3, 3, 4, 4, 2, 5, 6, 7, 7, 6, 5)))

 [1] 1 1 1 2 2 2 2 1 1 2 3 3 2 1
```

`inner_all` *inner_all*

Description

Allow to apply inner join on n dataframes, datatables, tibble

Usage

```
inner_all(..., keep_val = FALSE, id_v)
```

Arguments

`...`	are all the dataframes etc
`keep_val`	is if you want to keep the id column
`id_v`	is the common id of all the dataframes etc

Examples

```
datf1 <- data.frame(
        "id1"=c(1:5),
        "var1"=c("oui", "oui", "oui", "non", "non")
)

datf2 <- data.frame(
        "id1"=c(1, 2, 3, 7, 9),
        "var1"=c("oui2", "oui2", "oui2", "non2", "non2")
)

print(inner_all(datf1, datf2, keep_val=FALSE, id_v="id1"))

id1 var1.x var1.y
1   1    oui   oui2
2   2    oui   oui2
3   3    oui   oui2
```

intersect_all *intersect_all*

Description

Allows to calculate the intersection between n vectors

Usage

```
intersect_all(...)
```

Arguments

`...` is all the vector you want to calculate the intersection from

Examples

```
print(intersect_all(c(1:5), c(1, 2, 3, 6), c(1:4)))

[1] 1 2 3
```

join_n_lvl *join_n_lvl*

Description

Allow to see the progress of the multi-level joins of the different variables modalities. Here, multi-level joins is a type of join that usually needs a concatenation of two or more variables to make a key. But here, there is no need to proceed to a concatenation. See examples.

Usage

```
join_n_lvl(frst_datf, scd_datf, join_type = c(), lst_pair = list())
```

Arguments

`frst_datf` is the first data.frame (table)

`scd_datf` is the second data.frame (table)

`join_type` is a vector containing all the join type ("left", "inner", "right") for each variable

`lst_pair` is a lis of vectors. The vectors refers to a multi-level join. Each vector should have a length of 1. Each vector should have a name. Its name refers to the column name of multi-level variable and its value refers to the column name of the join variable.

Examples

```
datf3 <- data.frame("vil"=c("one", "one", "one", "two", "two", "two"),
                     "charac"=c(1, 2, 2, 1, 2, 2),
                     "rev"=c(1250, 1430, 970, 1630, 2231, 1875),
                     "vil2" = c("one", "one", "one", "two", "two", "two"),
                     "idl2" = c(1:6))
datf4 <- data.frame("vil"=c("one", "one", "one", "two", "two", "three"),
                    "charac"=c(1, 2, 2, 1, 1, 2),
                     "rev"=c(1.250, 1430, 970, 1630, 593, 456),
                     "vil2" = c("one", "one", "one", "two", "two", "two"),
                     "idl2" = c(2, 3, 1, 5, 5, 5))

print(join_n_lvl(frst_datf=datf3, scd_datf=datf4, lst_pair=list(c("charac" = "vil"), c("v
                 join_type=c("inner", "left")))

[1] "pair: charac vil"
|  |   0%
1
|= |  50%
2
|==| 100%
[1] "pair: vil2 idl2"
|  |   0%
one
|= |  50%
two
|==| 100%

  main_id.x vil.x charac.x rev.x vil2.x idl2.x main_id.y vil.y charac.y rev.y
1  1oneone1   one        1  1250    one      1      <NA>  <NA>       NA    NA
2  2oneone2   one        2  1430    one      2      <NA>  <NA>       NA    NA
3  2oneone3   one        2   970    one      3  2oneone3   one        2  1430
4  1twotwo4   two        1  1630    two      4      <NA>  <NA>       NA    NA
  vil2.y idl2.y
1   <NA>     NA
2   <NA>     NA
3    one      3
4   <NA>     NA
```

left_all *left_all*

Description

Allow to apply left join on n dataframes, datatables, tibble

Usage

```
left_all(..., keep_val = FALSE, id_v)
```

Arguments

`...`	are all the dataframes etc
`keep_val`	is if you want to keep the id column
`id_v`	is the common id of all the dataframes etc

Examples

```
datf1 <- data.frame(
        "id1"=c(1:5),
        "var1"=c("oui", "oui", "oui", "non", "non")
)

datf2 <- data.frame(
        "id1"=c(1, 2, 3, 7, 9),
        "var1"=c("oui2", "oui2", "oui2", "non2", "non2")
)

print(left_all(datf1, datf2, datf2, datf2, keep_val=FALSE, id_v="id1"))

  id1 var1.x var1.y var1.x.x var1.y.y
1   1    oui   oui2     oui2     oui2
2   2    oui   oui2     oui2     oui2
3   3    oui   oui2     oui2     oui2
4   4    non   <NA>     <NA>     <NA>
5   5    non   <NA>     <NA>     <NA>#'
print(left_all(datf1, datf2, datf2, keep_val=FALSE, id_v="id1"))

  id1 var1.x var1.y var1
1   1    oui   oui2 oui2
2   2    oui   oui2 oui2
3   3    oui   oui2 oui2
4   4    non   <NA> <NA>
5   5    non   <NA> <NA>
```

`pairs_findr` *pairs_findr*

Description

Takes a character as input and detect the pairs of pattern, like the parenthesis pais if the pattern is "(" and then ")"

Usage

```
pairs_findr(inpt, ptrn1 = "(", ptrn2 = ")")
```

Arguments

`inpt`	is the input character
`ptrn1`	is the first pattern ecountered in the pair
`ptrn2`	is the second pattern in the pair

Examples

```
print(pairs_findr(inpt="ze+(yu*45/(jk+zz)*(o()p))-(re*(rt+qs)-fg)"))

[[1]]
 [1] 4 1 1 3 2 2 3 4 6 5 5 6

[[2]]
 [1]  4 11 17 19 21 22 24 25 27 31 37 41
```

pairs_findr_merger *pairs_findr_merger*

Description

Takes two different outputs from pairs_findr and merge them. Can be usefull when the pairs consists in different patterns, for example one output from the pairs_findr function with ptrn1 = "(" and ptrn2 = ")", and a second output from the pairs_findr function with ptrn1 = "" and ptrn2 = "".

Usage

```
pairs_findr_merger(lst1 = list(), lst2 = list())
```

Arguments

lst1	is the first ouput from pairs findr function
lst2	is the second ouput from pairs findr function

Examples

```
print(pairs_findr_merger(lst1=list(c(1, 2, 3, 3, 2, 1), c(3, 4, 5, 7, 8, 9)),
                         lst2=list(c(1, 1), c(1, 2))))

[[1]]
[1] 1 1 2 3 4 4 3 2

[[2]]
[1] 1 2 3 4 5 7 8 9

print(pairs_findr_merger(lst1=list(c(1, 2, 3, 3, 2, 1), c(3, 4, 5, 7, 8, 9)),
                         lst2=list(c(1, 1), c(1, 11))))

[[1]]
[1] 1 2 3 4 4 3 2 1

[[2]]
[1]  1  3  4  5  7  8  9 11

print(pairs_findr_merger(lst1=list(c(1, 2, 3, 3, 2, 1), c(3, 4, 5, 8, 10, 11)),
                         lst2=list(c(4, 4), c(6, 7))))

[[1]]
[1] 1 2 3 4 4 3 2 1
```

```
[[2]]
[1]  3  4  5  6  7  8 10 11

print(pairs_findr_merger(lst1=list(c(1, 2, 3, 3, 2, 1), c(3, 4, 5, 7, 10, 11)),
                         lst2=list(c(4, 4), c(8, 9))))

[[1]]
[1] 1 2 3 3 4 4 2 1

[[2]]
[1]  3  4  5  7  8  9 10 11

print(pairs_findr_merger(lst1=list(c(1, 2, 3, 3, 2, 1), c(3, 4, 5, 7, 10, 11)),
                         lst2=list(c(4, 4), c(18, 19))))

[[1]]
[1] 1 2 3 3 2 1 4 4

[[2]]
[1]  3  4  5  7 10 11 18 19

print(pairs_findr_merger(lst1 = list(c(1, 1, 2, 2, 3, 3), c(1, 25, 26, 32, 33, 38)),
                        lst2 = list(c(1, 1, 2, 2, 3, 3), c(7, 11, 13, 17, 19, 24))))

[[1]]
 [1] 1 2 2 3 3 4 4 1 5 5 6 6

[[2]]
 [1]  1  7 11 13 17 19 24 25 26 32 33 38

print(pairs_findr_merger(lst1 = list(c(1, 1, 2, 2, 3, 3), c(2, 7, 9, 10, 11, 15)),
                         lst2 = list(c(3, 2, 1, 1, 2, 3, 4, 4), c(1, 17, 18, 22, 23, 29,

[[1]]
 [1] 6 5 1 1 2 2 3 3 4 4 5 6 7 7

[[2]]
 [1]  1  2  7  9 10 11 15 17 18 22 23 29 35 40

print(pairs_findr_merger(lst1 = list(c(1, 1), c(22, 23)),
                         lst2 = list(c(1, 1, 2, 2), c(3, 21, 27, 32))))

[[1]]
[1] 1 1 2 2 3 3

[[2]]
[1]  3 21 22 23 27 32
```

pairs_insertr *pairs_insertr*

Description

Takes a character representing an arbitrary condition (like ReGeX for example) or an information (to a parser for example), vectors containing all the pair of pattern that potentially surrounds condition (flagged_pair_v and corr_v), and a vector containing all the conjuntion character, as input and returns the character with all or some of the condition surrounded by the pair characters. See examples. All the pair characters are inserted according to the closest pair they found priotizing those found next to the condition and on the same depth-level and , if not found, the pair found at the n+1 depth-level.

Usage

```
pairs_insertr(
  inpt,
  algo_used = c(1:3),
  flagged_pair_v = c(")", "]"),
  corr_v = c("(", "["),
  flagged_conj_v = c("&", "|")
)
```

Arguments

`inpt` is the input character representing an arbitrary condition, like ReGex for example, or information to a parser for example

`algo_used` is a vector containing one or more of the 3 algorythms used. The first algorythm will simply put the pair of parenthesis at the condition surrounded and/or after a character flagged (in flagged_conj_v) as a conjunction. The second algorythm will put parenthesis at the condition that are located after other conditions that are surrounded by a pair. The third algorythm will put a pair at all the condition, it is very powerfull but takes a longer time. See examples and make experience to see which combination of algorythm(s) is the most efficient for your use case.

`flagged_pair_v` is a vector containing all the first character of the pairs

`corr_v` is a vector containing all the last character of the pairs

`flagged_conj_v` is a vector containing all the conjunction character

Examples

```
print(pairs_insertr(inpt = "([one]|two|twob)three(four)", algo_used = c(1)))

[1] "([one]|[two]|[twob])three(four)"

print(pairs_insertr(inpt = "(one|[two]|twob)three(four)", algo_used = c(2)))

[1] "(one|[two]|[twob])(three)(four)"

print(pairs_insertr(inpt = "(oneA|[one]|two|twob)three(four)", algo_used = c(1, 2)))

[1] "(oneA|[one]|[two]|[twob])(three)(four)"

print(pairs_insertr(inpt = "(oneA|[one]|two|twob)three(four)", algo_used = c(1, 2, 3)))

[1] "([oneA]|[one]|[two]|[twob])(three)(four)"
```

```
print(pairs_insertr(inpt = "(oneA|[one]|two|twob)three(four)", algo_used = c(3)))

[1] "([oneA]|[one]|(two)|(twob))(three)(four)"

print(pairs_insertr(inpt = "(oneA|[one]|two|twob)three((four))", algo_used = c(3)))

[1] "([oneA]|[(one)]|(two)|(twob))(three)((four))"
```

`pairs_insertr2` *pairs_insertr2*

Description

Takes a character representing an arbitrary condition (like ReGeX for example) or an information (to a parser for example), vectors containing all the pair of pattern that potentially surrounds condition (flagged_pair_v and corr_v), and a vector containing all the conjuntion character, as input and returns the character with all or some of the condition surrounded by the pair characters. See examples. All the pair characters are inserted according to the closest pair they found priotizing those found next to the condition and on the same depth-level and , if not found, the pair found at the n+1 depth-level.

Usage

```
pairs_insertr2(
  inpt,
  algo_used = c(1:3),
  flagged_pair_v = c(")", "]"),
  corr_v = c("(", "["),
  flagged_conj_v = c("&", "|"),
  method = c("(", ")")
)
```

Arguments

`inpt` is the input character representing an arbitrary condition, like ReGex for example, or information to a parser for example

`algo_used` is a vector containing one or more of the 3 algorythms used. The first algorythm will simply put the pair of parenthesis at the condition surrounded and/or after a character flagged (in flagged_conj_v) as a conjunction. The second algorythm will put parenthesis at the condition that are located after other conditions that are surrounded by a pair. The third algorythm will put a pair at all the condition, it is very powerfull but takes a longer time. See examples and make experience to see which combination of algorythm(s) is the most efficient for your use case.

`flagged_pair_v` is a vector containing all the first character of the pairs

`corr_v` is a vector containing all the last character of the pairs

`flagged_conj_v` is a vector containing all the conjunction character

`method` is length 2 vector containing as a first index, the first character of the pair inserted, and at the last index, the second and last character of the pair

Examples

```
print(pairs_insertr2(inpt = "([one]|two|twob)three(four)", algo_used = c(1), method = c("

[1] "([one]|(two)|(twob))three(four)"

print(pairs_insertr2(inpt = "([one]|two|twob)three(four)", algo_used = c(1), method = c("

[1] "([one]|[two]|[twob])three(four)"

print(pairs_insertr2(inpt = "(oneA|[one]|two|twob)three(four)", algo_used = c(1, 2)))

[1] "(oneA|[one]|(two)|(twob))(three)(four)"

print(pairs_insertr2(inpt = "(oneA|[one]|two|twob)three(four)", algo_used = c(1, 2), meth
                     flagged_pair_v = c(")", "]", "#"), corr_v = c("(", "[", "-")))

[1] "(oneA|[one]|-two#|-twob#)-three#(four)"

print(pairs_insertr2(inpt = "(oneA|[one]|two|twob)three(four)", algo_used = c(1, 2, 3)))

[1] "((oneA)|[one]|(two)|(twob))(three)(four)"

print(pairs_insertr2(inpt = "(oneA|[one]|two|twob)three(four)", algo_used = c(3), method

[1] "([oneA]|[one]|[two]|[twob])[three](four)"

print(pairs_insertr2(inpt = "(oneA|[one]|two|twob)three((four))", algo_used = c(3)))

[1] "((oneA)|[one]|(two)|(twob))(three)((four))"
```

Index