DichotomousKey package insight

--AN OPEN SOURCE PACKAGE FOR CLASSIFICATION CODED BY HAO LI

Fitness for purpose

DICHOTOMOUS KEY TABLE CAN BE USEFUL FOR CLASSIFICATION, BUT OPERATION CAN BE TIME CONSUMING.

Background – characteristics of dk tables 1

- Can be found in publications, or made by researchers
- Once a table is made, it is hard to manually edit it and keep the index and reference consistent.
- Publications may keep copyrights of the dk tables, therefore the users may obtain more than 1, but not all dk tables at once
- Tables are on different levels, a more detailed table may be appended to a more general table to continue classification. This may be very time consuming during manual operation
- Contents of tables may be in multimedia form(e.g. audios, pics, videos, delocalized data…). Tables must be extensible in columns

Dk tables 2

▶ The potential use of dk tables may also include new species registration and phenotype extraction. Now we can simply list all phenotypes by back referencing. In the future, AI for natural language processing may help the searching.

An simplified dk table included in the package

```
dk_eg

##   id                  P             G ref
## 1  1      plants like       plants   2
## 2  1      animal like       animal   3
## 3  2      woody trunk         tree   0
## 4  2        soft stem       flower   0
## 5  3 live under water         fish   0
## 6  3     live on land land_animal   4
## 7  4             hair       mammal   0
## 8  4         feathers         bird   0
```

Brief

The existing form of tables is good.

But we want it to be automated...

so that the functions are highly extensible addressing the new requirements of researchers and applicable to manipulate the data in large scale.

Requirements

▶ It must enables identification(identify species with known phenotype) and extraction(extract phenotype data from know specie).

▶ It must enables the conversion between a table-like data structure to a list-like data structure so that visual display of the structure.

▶ Appending must be enabled with list editing, and then all table-like objects can be converted to list for appending and then convert back(because list-like object do not link by index and reference number).

Outcome

TO PROVE I AM NOT LYING, TAKE A LOOK AT THE FUNCTIONS THAT ARE CURRENTLY AVAILIABLE

1. dk_classify()

▶ You can start an interactive searching by calling dk_classify().

▶ You can let it take in your own dk table**(type ?dk_eg in the command line to get the data structure required to import your own dataset)**

▶ Now, classify a thing in the default table, and assigning it to the variable called 'myselflol'.

```
> myselflol = dk_classify(dk_eg, asp = 1:4)
classification start from dk_eg
 in interactive mode
 to abort type'q'
Phenotypical characteristic:
Key ::  1  ::
plants like :: || ::plants
Key ::  2  ::
animal like :: || ::animal
```

1. dk_classify()

- In this case I am animal like, live on land and have hair(type the corresponding key and hit enter)

- Oh, I belong to the type 'mammal'

```
animal like :: || ::animal
2
Phenotypical characteristic:
Key ::  1  ::
live under water :: || ::fish
Key ::  2  ::
live on land :: || ::land_animal
2
Phenotypical characteristic:
Key ::  1  ::
hair :: || ::mammal
Key ::  2  ::
feathers :: || ::bird
1
> myselflol
   id    P       G ref
7  4 hair mammal    0
```

2.dk_extract

▶ Given that we have a individual mammal, extract its phenotype.

```
dk_extract(dk_eg,Gtarget = 'mammal')

##                 P          G
## 2   animal like      animal
## 6 live on land land_animal
## 7         hair      mammal
```

3.dk_as_list()

▶ Given a table like key dataset, convert it to a list-like object

```
list_eg = dk_as_list(dk_eg)
str(list_eg)

## List of 3
##  $      :List of 4
##   ..$      :List of 2
##   .. ..$ pause: logi TRUE
##   .. ..$ prim :'data.frame': 1 obs. of  2 variables:
##   .. .. ..$ P: chr "woody trunk"
##   .. .. ..$ G: chr "tree"
##   ..$      :List of 2
##   .. ..$ pause: logi TRUE
……
```

4. List appending(example)

▶ Now the classification for bird is too general, if I get another table classifying 'big bird' and 'small bird', I can create a new list by appending, knowing that 'bird' in the first list is in list_eg[[2]][[2]][[2]]

```r
dk_bird = data.frame(id = c(1,1),P = c('Big','Small'),G=  c('Big
Bird','Small Bird'),ref = c(0,0),stringsAsFactors = F)
list_bird = dk_as_list(dk_bird)
```

```r
list_new = list_eg
list_bird$prim = list_new[[2]][[2]][[2]]$prim
list_new[[2]][[2]][[2]] = list_bird
list_new[[2]][[2]][[2]] = as.list(list_new[[2]][[2]][[2]])
```

5.list_as_dk()

▶ And convert it back to get a new table. Use it just like you use the table dk_eg

```
dk_new = list_as_dk(list_new)
dk_new

##       id  P                   G              ref
##  [1,] "1" "plants like"       "plants"       "2"
##  [2,] "1" "animal like"       "animal"       "3"
##  [3,] "2" "woody trunk"       "tree"         "0"
##  [4,] "2" "soft stem"         "flower"       "0"
##  [5,] "3" "live under water"  "fish"         "0"
##  [6,] "3" "live on land"      "land_animal"  "4"
##  [7,] "4" "hair"              "mammal"       "0"
##  [8,] "4" "feathers"          "bird"         "5"
##  [9,] "5" "Big"               "Big Bird"     "0"
## [10,] "5" "Small"             "Small Bird"   "0"
```

appended ←

To use your own data

▶ You can import worksheets from excel.

▶ Or save excel as .csv(comma separatd values), import it with r function read.csv()/write with write.csv(), type a ? In front of each function in the R command line to show help documents

▶ More information about reading/writing files in R can be found in a number of R books or online tutorials

Installation and loading

GET THE PACKAGE WORKING

Development version

- You need an R to start working.

- https://cran.r-project.org/index.html

- Cran is '**The Comprehensive R Archive Network**'

- R is for '**The R Project for Statistical Computing**'

- Then get devtools package, typing into the command or paste this:

- install.packages('devtools')

- Then install via GitHub

- devtools:install_github('HaoLi111/dichotomousKey')

Development & Bug report

WHERE TO FIND SUPPORT AND HOW TO COLLABORATE

Development & Bug report

- **Extension**

- Feel free to share this open source package. For future support or report bugs or writing feedbacks, please contact me by email or on GitHub.

- Thank you for your support!

- https://github.com/HaoLi111/dichotomousKey

Thanks for watching

Hao Li https://github.com/HaoLi111

Data structure and multimedia extention

IN DEVELOPMENT

Working with big data

IN DEVELOPMENT