

xlsReadWrite

Andrej Blejec

15. maj 2009

1 xlsReadWrite

Opozorilo:

Paket `xlsReadWrite` od verzije 2.9 naprej ni več v seznamu podprtih paketov. Najdete ga lahko še na CRAN v Arhivu.

Priporočljiva je kaka druga možnost, najboljša je najbrž uporaba RODBC, nekaj takega kot je opisano v razdelku 3.

Paket `xlsReadWrite` omogoča povezavo R in programa Excel v Excelovem domačem formatu. Na voljo sta osnovni funkciji `read.xls` in `write.xls`. To seveda zelo olajša izmenjavo tabel. To učinkovito nadomešča bolj nerodno možnost uporabe funkcij `write.table` in `read.table`, katere prednost pa je zapis v ASCII formatu.

Primer uporabe funkcij `read.xls` in `write.xls`:

```
> library(xlsReadWrite)
```

```
xlsReadWrite version 1.3.2 (Build 163)  
Copyright (C) 2007, Hans-Peter Suter, Treetron, Switzerland.
```

```
This package can be freely distributed and used for any  
purpose. It comes with ABSOLUTELY NO GUARANTEE at all.  
xlsReadWrite has been written with Delphi and contains  
code from a 3rd party library. Our own code is free (GPLv2).
```

```
Please refer to http://treetron.googlepages.com for feedback,  
bugreports, donations, updates and the xlsReadWritePro version.
```

```
> n <- 4  
> bla = data.frame(faktor = LETTERS[1:n], znak = letters[1:n],  
+   prva = 1:n, druga = (1:n/n), tretja = (1:n)/3, cetrt = rep(c(TRUE,  
+   FALSE), n/2))  
> dimnames(bla)[[1]] = paste("vrstica", 1:n)  
> bla
```

	faktor	znak	prva	druga	tretja	cetrtta
vrstica 1	A	a	1	0.25	0.3333333	TRUE
vrstica 2	B	b	2	0.50	0.6666667	FALSE
vrstica 3	C	c	3	0.75	1.0000000	TRUE
vrstica 4	D	d	4	1.00	1.3333333	FALSE

```
> str(bla)
```

```
'data.frame':      4 obs. of  6 variables:
 $ faktor: Factor w/ 4 levels "A","B","C","D": 1 2 3 4
 $ znak  : Factor w/ 4 levels "a","b","c","d": 1 2 3 4
 $ prva  : int  1 2 3 4
 $ druga : num  0.25 0.5 0.75 1
 $ tretja: num  0.333 0.667 1 1.333
 $ cetrtta: logi  TRUE FALSE TRUE FALSE
```

```
> write.xls(bla, file = "../data/bla.xls")
```

```
> read.xls(file = "../data/bla.xls")
```

	faktor	znak	prva	druga	tretja	cetrtta
vrstica 1	A	a	1	0.25	0.3333333	1
vrstica 2	B	b	2	0.50	0.6666667	0
vrstica 3	C	c	3	0.75	1.0000000	1
vrstica 4	D	d	4	1.00	1.3333333	0

Pika ali vejica, to ni več vprašanje!

- write.xls: zapis je tak, da se pravilno odpre ne glede na nastavev Excela
- read.xls: pravilno prebere, ne glede na to, ali je Excel nastavljen na decimalno piko ali vejico

Datoteka blav.xls je shranjena z vejico, v R pa se prenese brez napak.

```
> read.xls("../data/blav.xls")
```

	znak	prva	druga	tretja
vrstica 1	A	1	0.1	0.3333333
vrstica 2	B	2	0.2	0.6666667
vrstica 3	C	3	0.3	1.0000000
vrstica 4	D	4	0.4	1.3333333
vrstica 5	E	5	0.5	1.6666667
vrstica 6	F	6	0.6	2.0000000
vrstica 7	G	7	0.7	2.3333333
vrstica 8	H	8	0.8	2.6666667
vrstica 9	I	9	0.9	3.0000000
vrstica 10	J	10	1.0	3.3333333

2 Manjkajoče vrednosti

poglejmo še, kako se zapišejo manjkajoče vrednosti. V drugo vrstico vpišemo po vrsti izjemne vrednosti, jih zapišemo z `write.xls` in preberemo z `read.xls`:

- factor: ne zapise NA, javi napako
- character: ne zapise NA
- numeric: včasih ? NA zapise prazno celico, prebere ko 0 zapiše in prebere NA
- numeric: NaN zapise kot NaN in pravilno prebere kot NaN
- logical: zapise kot 1/0 ter NA in prebere TRUE/FALSE in NA

```
> bla[2, 3] <- NA
> bla[2, 4] <- NaN
> bla[2, 5] <- Inf
> bla[2, 6] <- NA
> bla
```

```
      faktor znak prva druga   tretja cetrt
vrstica 1      A  a   1  0.25 0.3333333  TRUE
vrstica 2      B  b  NA  NaN      Inf    NA
vrstica 3      C  c   3  0.75 1.0000000  TRUE
vrstica 4      D  d   4  1.00 1.3333333  FALSE
```

```
> str(bla)
```

```
'data.frame':      4 obs. of  6 variables:
 $ faktor: Factor w/ 4 levels "A","B","C","D": 1 2 3 4
 $ znak  : Factor w/ 4 levels "a","b","c","d": 1 2 3 4
 $ prva  : int   1 NA 3 4
 $ druga : num  0.25 NaN 0.75 1
 $ tretja: num  0.333 Inf 1 1.333
 $ cetrt : logi  TRUE NA TRUE FALSE
```

```
> write.xls(bla, file = "../data/bla2.xls")
> ble <- read.xls("../data/bla2.xls")
> ble
```

```
      faktor znak prva druga   tretja cetrt
vrstica 1      A  a   1  0.25 0.3333333    1
vrstica 2      B  b  NA  NaN      Inf    NA
vrstica 3      C  c   3  0.75 1.0000000    1
vrstica 4      D  d   4  1.00 1.3333333    0
```

```
> str(ble)
```

```
'data.frame':      4 obs. of  6 variables:
 $ faktor: Factor w/ 4 levels "A","B","C","D": 1 2 3 4
 $ znak  : Factor w/ 4 levels "a","b","c","d": 1 2 3 4
 $ prva  : num  1 NA 3 4
 $ druga : num  0.25 NaN 0.75 1
 $ tretja: num  0.333 Inf 1 1.333
 $ cetrt: num  1 NA 1 0
```

Pri branju z `read.xls` je priporočljiva navedba razredov spremenljivk, z navedbo pravega razreda prebere

- character,factor: prebere pravilno
- numeric: Prazno prebere kot NA?
- numeric: NaN prebere kot NaN, Inf bere kot Inf
- logical: 0,1 prebere kot FALSE, TRUE; prazno bere kot FALSE

```
> ble <- read.xls("../data/bla2.xls", colClasses = c("character",
+ "factor", "character", "integer", "double", "double",
+ "logical"))
> ble
```

	faktor	znak	prva	druga	tretja	cetrt
vrstica 1	A	a	1	0.25	0.3333333	TRUE
vrstica 2	B	b	NA	NaN	Inf	FALSE
vrstica 3	C	c	3	0.75	1.0000000	TRUE
vrstica 4	D	d	4	1.00	1.3333333	FALSE

```
> str(ble)
```

```
'data.frame':      4 obs. of  6 variables:
 $ faktor: Factor w/ 4 levels "A","B","C","D": 1 2 3 4
 $ znak  : chr  "a" "b" "c" "d"
 $ prva  : int  1 NA 3 4
 $ druga : num  0.25 NaN 0.75 1
 $ tretja: num  0.333 Inf 1 1.333
 $ cetrt: logi  TRUE FALSE TRUE FALSE
```

Skratka, z manjkajočimi vrednostmi je treba biti pazljiv!!!

3 RODBC

Druga možnost je morda tale uporaba RODBC paketa, ki deluje dobro. Opazil sem, da argument `rownames` ne deluje tako, kot je opisano v pomoči oz. da pričakuje, da bo prvi stolpec poimenovan "rownames".

```
> library(RODBC)
> read.xls <- function(file, ..., sheet = "Sheet1", cch = "#",
+   comment = FALSE, rownames = TRUE, colnames = TRUE) {
+   z <- odbcConnectExcel(file)
+   myframe <- sqlFetch(z, sheet, rownames = rownames,
+     colnames = !colnames, ...)
+   close(z)
+   rownames(myframe) <- myframe[, 1]
+   myframe <- myframe[, -1]
+   commentLines <- grep(paste("^", cch, sep = ""), rownames(myframe))
+   if (!is.null(commentLines) & length(commentLines) >
+     0 & !comment)
+     myframe <- myframe[-commentLines, ]
+   if (comment)
+     myframe <- rownames(myframe[commentLines, ])
+   invisible(myframe)
+ }
```

Preberimo podatke

```
> data <- read.xls("../data/bla2.xls")
> data
```

	faktor	znak	prva	druga	tretja	cetrta
vrstica 1	A	a	1	0.25	0.3333333	1
vrstica 2	B	b	NA	NaN	Inf	NA
vrstica 3	C	c	3	0.75	1.0000000	1
vrstica 4	D	d	4	1.00	1.3333333	0

Struktura podatkov

```
> str(data)

'data.frame':      4 obs. of  6 variables:
 $ faktor: Factor w/ 4 levels "A","B","C","D": 1 2 3 4
 $ znak  : Factor w/ 4 levels "a","b","c","d": 1 2 3 4
 $ prva  : num  1 NA 3 4
 $ druga : num  0.25 NaN 0.75 1
 $ tretja: num  0.333 Inf 1 1.333
 $ cetrta: num  1 NA 1 0
```

Branje z drugega lista

```
> (read.xls("../data/bla3.xls", sheet = "Sheet2"))
```

```
      faktor znak
vrstica 1      A   a
vrstica 2      B   b
vrstica 4      D   d
```

Izpis dokumentacije zapisane v komentarjih

```
> (read.xls("../data/bla3.xls", sheet = "Sheet2", comment = TRUE))
```

```
[1] "# komentraciji so lahko od druge vrstice dalje"
[2] "#vrstica 3"
[3] "# komentar: vrstica 3 je izključena s komentarjem"
[4] "# opis podatkov:"
```

Glej: <http://tolstoy.newcastle.edu.au/R/help/04/12/8638.html>

4 Linux

Za Linux predlagajo uporabo perl:

<http://www-106.ibm.com/developerworks/linux/library/l-pexcel/>

(povezava ni preverjena)