

**“EXCELTOR” THE MODULE
FOR AN ANALYSIS OF
GEOBOTANICAL DATA BASED
ON THE EXCEL AND THE
STATISTICAL PACKAGE R**

Novakovskiy A.B.
Institute of biology
Komi Science Center

```
102 #library("vegan")
103
104 result <- tryCatch({
105   library("vegan")
106   library(dendextend)
107   img <- generatePathToImageFile(args$out_file)
108   tiff(img, width = Pictwidth, height = PictHeight, units = "px")
109   data_matrix=t(data_matrix) #we need to transpose matrix before using
110
111   # drawing
112   if(Type_of_input == 0){ #calculate dissimilarity matrix
113     if (index == "pearson" || index == "kendall") {
114       dist = as.dist((1-cor(data_matrix, method = index))/2);
115       capt = paste("Index: (1 - '", index, "' correlation)/2. Agglomerative method: '", agglom, "'");
116     } else
117     {
118       dist=vegdist(data_matrix, method=index, binary=is_binary);
119       capt = paste("Index: '", index, "'. Agglomerative method: '", agglom, "'");
120     }
121   }
122   if (Type_of_input == 1) {
123     # Do nothing, it is a distance matrix
124     capt = paste("Distance matrix. Agglomerative method: '", agglom, "'");
125     dist = as.dist(data_matrix)
126
```

Principal scheme of ExcelToR

	A	B	C	D	E	F	G
1	точка	горизонт	фракция	Cu	Pb	Zn	Ni
2	E-EL	OB1	total	8,1	13	32	19
3	E-EL	OB1	con	6,6	67	34	21
4	E-EL	B	total	8,1	13	32	19
5	E-EL	B	con	15	56	38	32
6	E-GC	ABg	total	5,5	14	31	14
7	E-GC	ABg	con	14,3	33	41	30
8	E-GC	Bg	total	8,3	12	41	18
9	E-GC	Bg	con	6	71	34	19
10	E-GC	B1	total	9	11,7	36	19
11	E-GC	B1	con	8,8	67	35	24

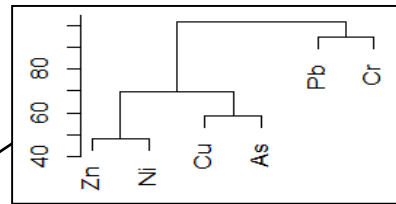
Excel source file

XML file:
Source data,
captions,
settings, etc.

```

39
40 # 1. parse XML
41 #####
42 # initial parsing
43 xml <- parsexml(args$in_file,args$out_file)
44
45 # get captions
46 captions <- getStringVectorFromXML(xml,"Captions",args$out_file)
47
48 # get main (data) matrix
49 data_matrix <- t(getNumericMatrixFromXML(xml,"DataMatrix",args$out_file))
50
51 # get configurations
52 cfg <- getElementFromXML(xml,"cfg",args$out_file)
53 index <- getStringValueFromXML(cfg,"similarityIndex",args$out_file)
54 is_binary <- getStringValueFromXML(cfg,"isBinaryIndex",args$out_file)
55 agglom <- getStringValueFromXML(cfg,"AgglomerativeMethod",args$out_file)
56
57

```



Result diagrams

Calculation in the "R"

	A	B	C	D	E	F	G
1	точка	горизонт	фракция	Cu	Pb	Zn	Ni
2	E-EL	OB1					19
3	E-EL	OB1					21
4	E-EL	B					19
5	E-EL	B					32
6	E-GC	ABg					14
7	E-GC	ABg					30
8	E-GC	Bg					18
9	E-GC	Bg					19
10	E-GC	B1	total	9	11,7	36	19
11	E-GC	B1	con	8,8	67	35	24

Final results

(diagrams and/or tables)

XML file:
result tables

Designed algorithms

- Cluster analysis
 - Single-linkage clustering
 - Complete-linkage clustering
 - Ward's method
 - UPGMA
- Ordination
 - PCA
 - NMS
- Heat Maps

Conclusion

- This add-on is easy to use and expands capabilities of Excel in the statistical analysis
- This approach of data communication between Excel and R allows to implement new algorithms of data analysis fast and easy

ib.komisc.ru/exceltor