



Математички факултет, Универзитет у Београду

## Упутство за Семинарски рад из предмета: Истраживање података 2

Аутори:  
Марко Никитовић  
Бошко Андрић

Предметни наставник:  
Проф. др Ненад Митић

Школска година 2023/2024

Комплетан пројекат се налази на: "<https://github.com/shotinform/proteinclustering>". Он укључује, поред овог и остала 2 документа (један је сам рад, док је други документ са свим добијеним сликама), све "Python" и "R" кодове који се могу покренути и директоријуме у којима се налазе генерисане слике и текстуални документи (у сваком поглављу главног рада је дата путања до релевантних слика, кодова и текстуалним документима), такође се могу пронаћи и подаци који су били дати на почетку.

## 1 Потребне библиотеке за програмски језик "Python"

Команде за преузимање свих потребних библиотека:

1. `pip install numpy`
2. `pip install matplotlib`
3. `pip install matplotlib_venn`
4. `pip install py4cytoscape`
5. `pip install pandas`
6. `pip install biopython`
7. `pip install networkx`
8. `pip install seaborn`

## 2 Потребне библиотеке за програмски језик "R"

1. `VennDiagram`
2. `clusterProfiler`
3. `org.Hs.eg.db`
4. `ReactomePA`
5. `pheatmap`
6. `grid`
7. `gridExtra`
8. `igraph`
9. `ggraph`
10. `dplyr`

Наведени пакети су неопходни да би се све скрипте у језику "R" могле покренути. Покретањем наредне скрипте можете преузети све потребне пакете:

```
1 install_if_missing <- function(pkg) {  
2   if (!requireNamespace(pkg, quietly = TRUE)) {  
3     install.packages(pkg)  
4   }  
5 }  
6  
7 cran_packages <- c("VennDiagram", "pheatmap", "gridExtra", "igraph", "  
8   gggraph", "dplyr")  
9  
10  apply(cran_packages, install_if_missing)  
11  
12  install_if_missing("BiocManager")  
13  
14  bioc_packages <- c("clusterProfiler", "org.Hs.eg.db", "ReactomePA")  
15  BiocManager::install(bioc_packages)  
16  
17  lapply(c(cran_packages, bioc_packages), library, character.only = TRUE)
```