

# Projekt 1 faza 2

Grupa CT

10 kwietnia 2016

Najpierw wczytamy dane oraz załączymy biblioteki, niezbędne do analizy danych.

```
setwd("~/Uczelnia/R i Big Data/CNK/")

library(dplyr)
library(RSQLite)

db <- dbConnect(SQLite(), "dane/dane.db")
dbListTables(db)
dane <- dbReadTable(db, "interakcje")
dbDisconnect(db)

dane$poczek <- strptime(dane$poczek, '%Y-%m-%d %H:%M:%S')
```

Sortujemy w grupach wyznaczonych przez id zwiedzającego oraz daty interakcji względem początku interakcji.

```
dane %>% group_by(dzien = as.Date(poczek), gosc) %>%
  arrange(poczek) -> dane

eks <- unique(dane$eksponat)
str(dane)

n <- nrow(dane)

# ostatni eksponat w scieżce
dane$ostatni <- c(dane$gosc[-n] != dane$gosc[-1], 1)
```

Zliczamy dla każdej pary eksponatów ( $eks_p_i, eks_p_j$ ) ile razy  $eks_p_i$  był w scieżce po  $eks_p_j$ .

```
tabela <- table(dane$eksponat[-n], dane$eksponat[-1],
               dane$ostatni[-n])
tabela <- tabela[,1]
```

Zliczamy ile łącznie razy występował dany eksponat we wszystkich scieżkach oraz liczymy prawdopodobieństwo przejścia od  $eks_p_i$  do  $eks_p_j$ .

```
licznosci <- table(dane$eksponat, dane$ostatni)[,1]
pstwa <- tabela / licznosci
save(pstwa, file = "pstwa.rda")
```

## Aplikacja

Na podstawie danych o prawdopodobieństwach przejść zrobiliśmy za pomocą pakietu *Shiny* aplikację, która ma możliwości takie jak:

- wykres słupkowy prawdopodobieństw
- wizualizacje najbliższych popularnych eksponatów
- wizualizacje ścieżki na mapie eksponatów.

## WNIOSKI

Na podstawie analiz za pomocą aplikacji doszliśmy do następujących wniosków:

- większość eksponatów jest odwiedzana kilka razy pod rząd
- największe prawdopodobieństwo przejścia występuje dla eksponatów, które są najmniej oddalone od eksponatu, w którym jesteśmy w danym momencie.
- zauważyliśmy, że odwiedzający wystawę poruszają się po niej zgodnie z ruchem wskazówek zegara.