

# Aspekt inżynierski

## Instrukcja instalacji i konfiguracji narzędzia

Mimkatz={Magda Waśniowska, Katarzyna Herba, Marcin Kania}  
projekt realizowany w ramach Języki i Narzędzia Programowania 2 - grupa R

14 czerwca 2016

### 1 Instrukcja aktualizacji danych

By odtworzyć dane dotyczące ocen i ankiet niezbędne do poprawnego działania aplikacji trzeba:

1. pobrać z bazy danych
2. przepuścić przez program napisany w  $C++$
3. przepuścić przez program napisany w  $R$

W tej części opiszemy, jak te dane pobrać i w jaki sposób je przeczytać.

#### Pobranie danych z bazy danych

W celu pobrania danych z bazy danych należy najpierw połączyć się z bazą danych. Można to zrobić z poziomu terminalu (będąc na komputerze podłączonym do sieci wydziałowej) za pomocą następującej komendy:

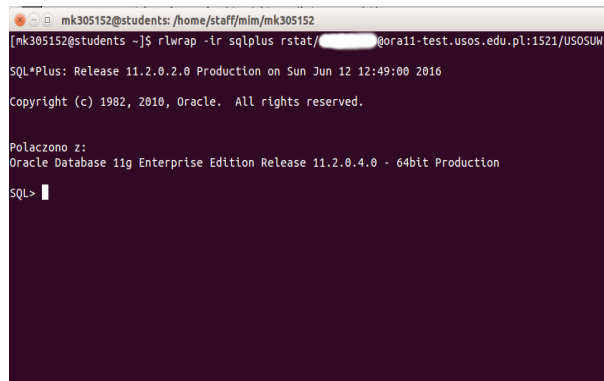
```
rlwrap -ir sqlplus USER/PASSWORD@HOST:PORT/SERVICE_NAME
```

- *USER* to nazwa użytkownika,
- *PASSWORD* to hasło,
- *HOST* to nazwa hosta,
- *PORT* to numer portu,
- *SERVICE\_NAME* to nazwa usługi.

Powyższa komenda wymaga zainstalowanego na maszynie użytkownik terminalowego klienta baz danych Oracle - SQL\*Plus. Przykładowy widok uzyskiwania połączenia został przedstawiony na rysunku 1.

Będąc już połączonym z bazą danych należy wpisać następujące zapytania do niej:

- w celu pobrania danych dotyczących ocen



Rysunek 1: Przykładowe połączenie z bazą danych (hasło wymazane na białym)

```
set linesize 1000;
spool SCIEZKA_DLA_OCEN;
select KOD, PRZ_NAZWA, CDYD_KOD, NUMER_TERMINU, OCENA, OSOBA
from rstat_oceny
where jed_org_kod_biorca=10000000;
spool off;
```

- w celu pobrania danych dotyczących ankiet

```
set linesize 1000;
spool SCIEZKA_DLA_ANKIET;
select KOD, PRZ_NAZWA, ZAJ_KOD, PROWADZACY, WARTOSC_ODPOWIEDZI,
LICZBA_ODPOWIEDZI, TRESC_PYTANIA, CDYD_KOD
from rstat_czestosci_ankiet
where KOD like '%1000-%';
spool off;
```

Należy w nich podmienić *SCIEZKA\_DLA\_OCEN* na absolutną ścieżkę do pliku (czyli np. *home/dane/raw/oceny\_raw.txt*), do którego chcemy wgrać wyniki dla ocen. Analogicznie robimy z *SCIEZKA\_DLA\_ANKIET*.

Alternatywnie można uruchomić skrypt *data.sql*. Należy upewnić się, że skrypt znajduje się w katalogu, z którego łączymy się z bazą danych, a następnie wywołuje się go za pomocą polecenia:

```
@data.sql;
```

## Wstępne oczyszczenie danych za pomocą *jnpcleaner*

Dane wyeksportowane z bazy danych nie są w postaci, którą można bezpośrednio zaimportować do programu R. Jest w nich wiele białych znaków i różnych znaków przystankowych uniemożliwiających podział danych na odpowiednią liczbę kolumn. Program *jnpcleaner* służy do sparsowania wyjścia z bazy danych do formatu, który R może bezproblemowo odczytać.

### Instrukcja kompilacji *jnpcleaner*

Jeśli użytkownik nie posiada pliku wykonywalnego *jnpcleaner*, to powinien skompilować go używając źródeł. W tym celu należy:

1. otworzyć terminal,
2. przejść do katalogu zawierającego kod źródłowy *src* aplikacji *jnpcleaner*,
3. wywołać w terminalu z poziomu wyżej wymienionego katalogu polecenie *cmake...*,
4. wywołać teraz *make*.

W przypadku poprawnego wykonaniu powyższych czynności w katalogu *src* pojawi się plik wykonywalny *jnpcleaner*.

Aby użytkownik mógł skompilować *jnpcleaner* zgodnie z wyższą metodą, musi posiadać odpowiednie narzędzie tzn. kompilator oraz program *cmake*. Użytkownicy systemów z rodziny *UNIX* nie powinni mieć problemu z tym, gdyż zazwyczaj kompilator taki jest udostępniony wraz z dystrybucją systemu, więc muszą jedynie pobrać z sieci *cmake* (jest darmowy).

W przypadku użytkowników systemu *Windows*, polecamy zainstalować pakiet *Cygwin* - za jego pomocą można zainstalować kompilator *C++*. Sugerowany kompilator to *gcc* w wersji 4.9.2. Również, jak w przypadku systemów rodziny *UNIX*, należy pobrać program *cmake*.

### Instrukcja używania *jnpcleaner*

Program *jnpcleaner* uruchamia się komendą

```
./jnpcleaner
```

z poziomu terminalu. Aby program poprawnie zadziałał należy mu podać następujące argumenty:

- jeśli przetwarzamy dane dotyczące ocen i ankiet

```
./jnpcleaner "both" "GradesInPath" "GradesOutPath"  
"QuestionsInPath" "QuestionsOutPath";
```

- jeśli przetwarzamy tylko dane dotyczące ocen

```
./jnpcleaner "grades" "GradesInPath" "GradesOutPath";
```

- jeśli przetwarzamy tylko dane dotyczące ankiet

```
./jnpcleaner "questions" "QuestionsInPath" "QuestionsOutPath";
```

gdzie *GradesInPath*, *GradesOutPath*, *QuestionsInPath* i *QuestionsOutPath* to zmienne, za które należy podstawić absolutne ścieżki odpowiednio do:

- pliku zawierającego surowe dane z bazy danych dotyczące ocen,
- pliku (niekoniecznie utworzonego) mającego zawierać oczyszczone dane dotyczące ocen,
- pliku zawierającego surowe dane z bazy danych dotyczące ankiet,

- pliku (niekoniecznie utworzonego) mającego zawierać oczyszczone dane dotyczące ankiet.

Przykładowe wywołanie to:

```
./jnpcleaner "both" "/home/dane/grades_raw.txt" "/home/dane/grades_clean.txt"
"/home/dane/questions_raw.txt" "/home/dane/questions_clean.txt"
```

Na systemach typu *Windows* za terminal służy np. *cmd.exe*, wówczas zamiast

```
./jnpcleaner
```

należy użyć komendy

```
jnpcleaner.exe
```

Alternatywnie można wywołać plik bashowy *clean.sh*, po wcześniejszym podmienieniu ścieżek docelowych w nim. Nomenklatura zmiennych dla ścieżek jest dokładnie taka sama, jak wyżej. Po poprawnym podmienieniu *clean.sh* wywołuje się komendą *./clean.sh*. W przypadku systemu *Windows* należy skorzystać w udostępnionego w ramach środowiska *Cygwin* terminalu i w nim wywołać skrypt.

## Wtórne oczyszczenie danych w *R*

Program *oczyszczenie\_danych.R* służy do:

- zastąpienia encji zawierających oceny uzyskane przez studenta w różnych terminach (z tego samego przedmiotu) przez jedną encję z ostateczną oceną,
- usunięcia danych nieużywanych w aplikacji (oceny z przedmiotów prowadzonych poza MIM UW),
- wyfiltrowania roczników od semestru letniego 2008 włącznie,
- rozdzielenia przedmiotów, które mają taką samą nazwę ale inny kod,
- usunięcia namiarowych białych znaków występujących w danych,
- usunięcia danych o przedmiotach o których wiadomo, że nie są już prowadzone przez MIM UW.

Aby użyć programu, należy otworzyć go w programie *RStudio* i pobrać bibliotekę *plyr* poleceniem `install.packages("plyr")`.

Następnie należy w pliku uzupełnić ścieżki do plików, znajdujące się w pierwszych liniach programu:

```
path_read
```

to ścieżka, do danych uzyskanych po przepuszczeniu danych pobranych z bazy danych przez program *jnpcleaner*,

```
path_write
```

to ścieżka, pod którą zostaną zapisane uzyskane w ten sposób dane do aplikacji. Po ustawieniu ścieżek należy zaznaczyć cały kod *oczyszczenie\_danych.R* i uruchomić go komendą *Run*, znajdującą się w prawym górnym rogu aplikacji.

## 2 Instrukcja instalacji *R* i konfiguracji aplikacji *Shiny*

W celu uruchomienia naszej aplikacji trzeba posiadać odpowiednią wersję środowiska programowania *R* oraz stosowne biblioteki. W tej części podamy wymagane programy, biblioteki oraz opiszemy sposób konfiguracji aplikacji, by działała na maszynie użytkownika. Sugerujemy, aby użytkownik zainstalował program *RStudio*, gdyż jest to bardzo dobre środowisko do pracy w języku *R*.

1. Zainstalować program *R* (wersja minimum 3.3.0).
2. W programie *R* zainstalować następujące biblioteki:

```
library(shiny)
library(ggplot2)
library(plyr)
library(xtable)
library(reshape2)
```

3. W obu plikach **server.R** oraz **ui.R** uzupełnić ścieżki prowadzące do danych (pierwsze linie kodu). Przykładowo:

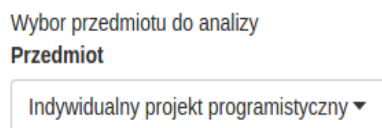
```
ankiety_path <- "home/user/Desktop/ankiety_clean.txt"
oceny_path <- "/home/user/Desktop/oceny.csv"
```

Warto tu zauważyć, że ścieżka *oceny\_path* (w plikach **server.R** i **ui.R**) musi być taka sama jak *path\_write* (w pliku *oczyszczenie\_danych.R*).

## 3 Instrukcja obsługi

Aplikacja pozwala na porównywanie udziału poszczególnych ocen w końcowych wynikach z danego przedmiotu na przestrzeni ostatnich lat. Prezentowane oceny są ostatecznymi jakie otrzymała dana osoba tzn. jeśli student *X* miał z przedmiotu *Y* oceny w terminach 1, 2 i 3 to w aplikacji została wykorzystana tylko ocena z terminu numer 3.

- W pierwszym kroku użytkownik wybiera przedmiot do analizy dostępny spośród rozwijanej listy przedmiotów prowadzonych przez jednostkę MIM UW (2).



Rysunek 2: Zrzut fragmentu ekranu prezentujący wybór przedmiotu

- W drugim kroku należy wskazać cykl/e dydaktyczne, których będą dotyczyły prezentowane dane (3). Dodatkowo użytkownik ma do wyboru dwa przyciski "Zaznacz wszystkie cykle" i "Odnznacz wszystkie cykle", które odpowiednio zaznaczają i odznaczają wszystkie możliwe do

wyboru cykle dydaktyczne. Dodatkowo użytkownik będzie miał do wyboru tylko te lata w których prowadzony był przedmiot wybrany w kroku numer 1.

**Cykl dydaktyczny**

☐ 2008L ☐ 2009L ☐ 2010L  
☐ 2011L ☐ 2012L ☐ 2013L  
☐ 2014L

Zaznacz wszystkie cykle

Odznacz wszystkie cykle

Rysunek 3: Zrzut fragmentu ekranu prezentujący wybór cyklu dydaktycznego

- Prezentacja wykresów odbywa się w dwa różne sposoby albo na jednym (7) albo w podziale na cykle (8). Opcja wyboru wykresu znajduje się w panelu sterowania po lewej stronie aplikacji (4).

☐ Czy pokazać na jednym wykresie, czy na osobnych?

Rysunek 4: Zrzut fragmentu ekranu prezentujący pytanie dotyczące prezentacji wykresu

- Użytkownik ma do wyboru cztery różne zakładki (5). Pierwsza z nich domyślnie włączona przedstawia wyżej omówione wykresy. Druga "Ankiety-Wykres" prezentuje wykres dotyczący danych z ankiet (9). Dwie kolejne są to zakładki z tabelami, pierwsza (10) zawiera tabelę z danymi dotyczącymi ocen natomiast druga (11) - ankiet. Dane prezentowane w tabelach są tymi samymi danymi, które są użyte do wygenerowania wykresów.

**Wybor wynikowego wykresu**

Oceny-Wykres   Ankiety-Wykres   Oceny-Tabela   Ankiety-Tabela

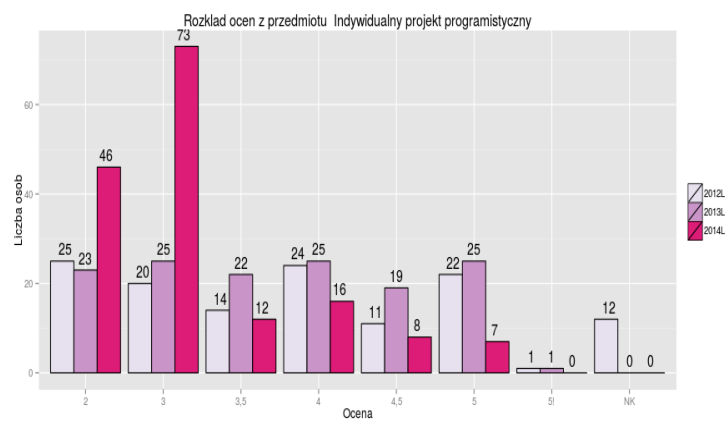
Rysunek 5: Zrzut fragmentu ekranu prezentujący dostępne zakładki

- Wykres prezentujący wyniki ankiet (9) dotyczy przedmiotu wybranego w kroku pierwszym (2) oraz cykli dydaktycznych (3) z kroku drugiego. Dodatkowo użytkownik powinien wybrać pytanie, dla którego chce zobaczyć wyniki. W panelu sterowania po lewej stronie służy do tego rozwijana lista (6).

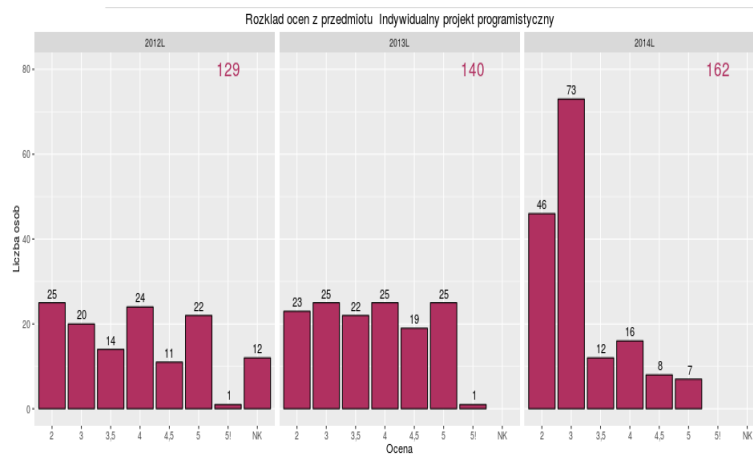
### Tresci pytan

Czy była mozliwosc prowadzenia notatek lub dostep do przygotowanych materialow?

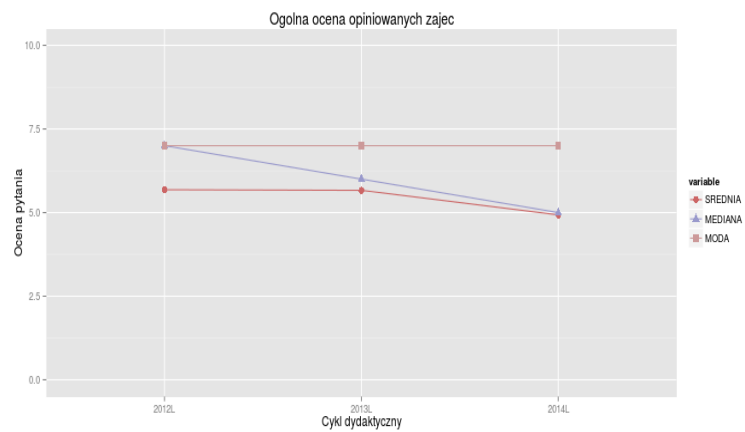
Rysunek 6: Zrzut fragmentu ekranu prezentujący wybór pytania ankiety



Rysunek 7: Wykres typu 1



Rysunek 8: Wykres typu 2



Rysunek 9: Wykres danych z ankiet

Show 10 entries Search:

Cykl dydaktyczny	2	3	3,5	4	4,5	5	5!	NK
2012L	25	20	14	24	11	22	1	12
2013L	23	25	22	25	19	25	1	0
2014L	46	73	12	16	8	7	0	0

Cykl dydaktyczny

Showing 1 to 3 of 3 entries Previous 1 Next

Rysunek 10: Tabela zestawienia ocen z przedmiotów na tle cykli dydaktycznych



Prowadzący	1	2	3	4	5	6	7
Artur Zaroda	0	0	0	2	0	1	0
Bartosz Szreder	0	1	0	3	3	1	9
Henryk Michalewski	0	0	1	2	1	1	1
Jakub Pawlewicz	2	0	0	4	5	6	12
Jakub Radoszewski	1	1	2	1	5	1	13
Lukasz Czajka	0	0	2	4	2	1	1
Lukasz Mazurek	1	0	0	0	2	7	2
Marcin Stefaniak	2	0	2	0	1	0	1
Michał Włodarczyk	0	0	1	1	1	1	2
Michał Zajac	0	0	0	2	0	0	0
Mirosław Michalski	0	0	0	0	0	0	2

Rysunek 11: Tabelarna zestawienie wyników konkretnego pytania z ankiet na tle prowadzących