

Prikaz socialnega omrežja - uporaba Apex, REST in R

red. prof. dr. Robert Leskovar

3. oktober 2021

Povzetek

V tej spominčici je prikazana uporaba podatkov o socialni mreži za izris grafa povezav med posamezniki. Demonstracijska spletna aplikacija, ki je izdelana z Oracle Application Express (APEX), omogoča dodajanje novih znancev v bazo podatkov. Del podatkov o socialnem omrežju (ime tistega, ki nekoga pozna, ime tistega, ki ga je nekdo prepoznal, moč poznanstva in datum zapisa) je dostopen preko spletnega programskega vmesnika REST (kratici REST API ali RESTful API, z besedo REpresentational State Transfer). V jeziku R je prikazan dostop do podatkov o socialnem omrežju preko zahteve HTTPS, obdelava prejetih podatkov v formatu JSON (odstranitev nebistvenih podatkov, oblikovanje podatkovnega okvira) ter izris socialnega omrežja.

Kazalo

1	Uvod	3
2	Baza podatkov, aplikacija in REST API	3
2.1	Baza podatkov	3
2.2	Aplikacija	4
2.3	Definicija REST API	4
3	Uporaba REST API v jeziku R	5
3.1	Pristop	5
3.2	Programska koda	6
4	Zaključek	7
	Literatura	8

Generirano 3. oktober 2021 18:37:30 CEST

1 Uvod

Le kompetentni ljudje lahko kompleksno tehnologijo uporabijo zavestno in koristno.

Robert Leskovar

Prikazan bo primer uporabe kompleksnih tehnologij:

- Oracle Application Express (kratica APEX) je orodje za malokodno programiranje.
- REpresentational State Transfer Application Programming Interface (kratica REST API ali RESTful API ali REST web API) je programski vmesnik, ki skladen z omejitvami arhitekturnega načina REST. Namen REST API je možnost povezovanja aplikacij, ki tečejo na različnih računalnikih. V tem primeru gre za instanco APEX-a v Oraclovem oblaku in poljubne delovne delovne postaje z omogočenim internetnim dostopom ter nameščenim jezikom R. Na APEX-u je nameščen program za zajem podatkov o socialnem omrežju ter definiran REST API v ORDS (Oracle REST Data Services, javanska aplikacija, ki omogoča definiranje REST API za bazo Oracle) za objavo dela podatkov o socialnem omrežju.
- R, RStudio, knjižnice *httr*, *jsonlite*, *igraph*, *dplyr*. R je jezik in okolje za statistične obdelave, vizualizacijo in predstavlja eno najpomembnejših orodij za strojno učenje in umetno inteligenco. RStudio je eden od uporabniških vmesnikov za jezik R. Omogoča, da program v jeziku R urejamo, izvajamo, pregledujemo rezultate v tekstovni in grafični obliki, poleg tega pa lahko priredimo nastavitve delovnega okolja.

Najprej bo prikazana baza podatkov ter aplikacija, sledi definiranje REST API. V jeziku R bo izveden klic REST API, podatki pa bodo shranjeni, obdelani ter prikazani v grafični obliki (graf povezav).

2 Baza podatkov, aplikacija in REST API

2.1 Baza podatkov

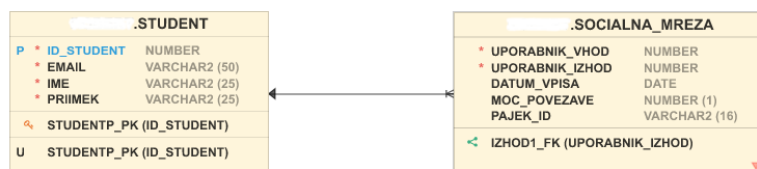
V bazi podatkov sta definirani tabeli STUDENT in SOCIALNA_MREZA ter povezani s tujim ključem (slika 1).

Navedeni definiciji tabel omogočata ponazoritev realnega stanja socialne mreže:

- en študent ima lahko več znancev
- en študent je lahko poznan več drugim študentom

S programom bomo zagotovili, da študent kot znance lahko vpiše le druge, ne pa sebe (UPORABNIK_VHOD je različen UPORABNIK_IZHOD). Sistem za upravljanje baze podatkov pa bo samodejno preprečeval podvojene vnose znancev (enaki pari UPORABNIK_VHOD,UPORABNIK_IZHOD).

Slika 1: Povezava dveh ključnih tabel.



2.2 Aplikacija

Aplikacija je razvita v orodju APEX. Demonstracijska spletna aplikacija zahteva prijavo. Navodila za prijavo so podana v spletni učilnici predmeta Računalništvo in informatika, prosojnice Uvodno predavanje na strani 5. Slika 2 prikazuje vstopno stran aplikacije po uspešni prijavi. Vnos znancev je preprost - pokli-

Slika 2: Vstopna stran aplikacije.

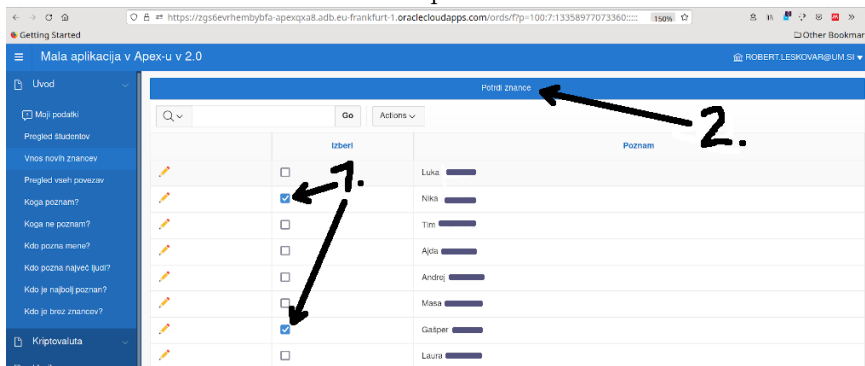


kamo potrditvena polja pred imeni in nato še gumb (Potrdi znance), ki za vsa označena poznanstva sproži zapis v bazo podatkov. Gumb se nahaja na vrhu in na dnu spletne strani (glej sliko 3). Na menujskem delu ekrana (trak na levi strani ekrana) lahko izvedemo še različne preglede znancev, neznancev, števila poznanstev in podobno. V tabeli SOCIALNA_MREZA so shranjeni podatki o socialni mreži. Baza podatkov sama preprečuje podvojene vnose, uporabniški vmesnik pa posamezniku ne omogoča vnosa poznanstva s samim seboj. Primer podatkov v tabeli je prikazan na sliki 4).

2.3 Definicija REST API

Administrator delovnega prostora najprej omogoči uporabo REST API za shemo (glej video [1]). Nato izbere SQL Workshop > RESTful Services in definira dostop. V konkretnem primeru gre za branje (ukaz SELECT v SQL) ter zahtevo GET v HTML. Ko je REST API definiran, lahko dobljeni URL prekopiramo v spletni brskalnik in pogledamo podatke v formatu JSON (slika 5). Kratica JSON

Slika 3: Vnos in potrditev znancev.



Slika 4: Primer podatkov o socialni mreži.

SOCIALNA_MREZA											
Table	Data	Indexes	Model	Constraints	Statistics	UI Defaults	Triggers	Dependencies	SQL	REST	Sample Queries
Query	Count Rows	Insert Row	Load Data								
EDIT	UPORABNIK_VHOD	UPORABNIK_IZHOD	DATUM_VPISA	MOC_POVEZAVE	PAJEK_ID						
<input type="checkbox"/>	20211	20980	09/30/2021	1	-						
<input type="checkbox"/>	20211	20212	09/30/2021	1	-						
<input type="checkbox"/>	20980	20211	09/30/2021	1	-						
<input type="checkbox"/>	20980	20212	09/30/2021	1	-						

pomeni JavaScript Object Notation. V tem formatu pogosto komunicirajo programi, ki npr. delujejo na različnih operacijskih sistemih in bazah podatkov ter pri izmenjavi uporabljajo protokol HTTP ali HTTPS - torej protokola za komunikacijo med spletnim strežnikom in spletnim brskalnikom.

Prvi del izpisa na sliki 5 so podatki, združeni pod oznako items (vir, ponor, utez, vpis), drugi del pa predstavljajo kontrolne informacije, ki jih na tem mestu ne bomo podrobneje obravnavali.

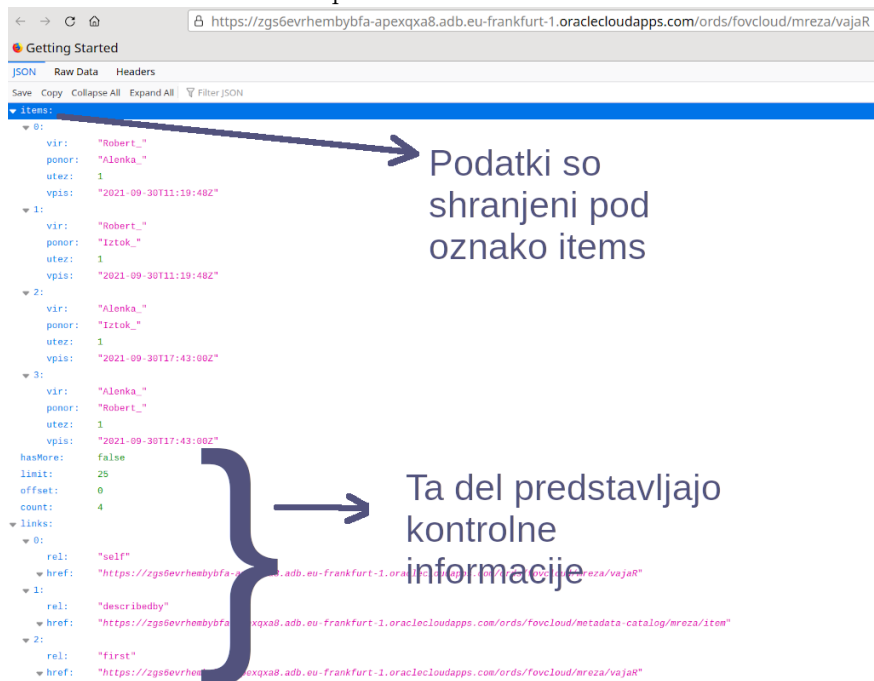
3 Uporaba REST API v jeziku R

3.1 Pristop

Napisati je potrebno programsko kodo v jeziku R, ki bo omogočala naslednje:

- nalaganje knjižnic za dostop do spletnih virov, obdelavo podatkov v formatu JSON, risanje grafov in manipulacijo s podatkovnimi okvirji v R (*httr*, *jsonlite*, *igraph*, *dplyr*)
- klic REST API in shranjevanje podatkov v spremenljivki *rezultat*
- pretvorba spremenljivke *rezultat* v spremenljivko *jsonRespParsed*, ki ima tip "List"

Slika 5: Prikaz podatkov REST API v brskalniku.



- pretvorba spremenljivke *jsonRespParsed* v spremenljivko *mreza*, ki ima tip "data frame"
- pretvorba podmnožice spremenljivk v objektu *mreza* v spremenljivko *net*, ki ima tip "graph.data.frame"
- izris grafa povezav v socialni mreži s pomočjo funkcije *cluster_edge_betweenness*

Pretvorbe so potrebne zato, ker moramo: a) vstopne podatke prečistiti (odstranimo kontrolne), b) pripraviti ustrezno strukturirane objekte in c) izbrati samo podatke za gradnjo grafa.

3.2 Programska koda

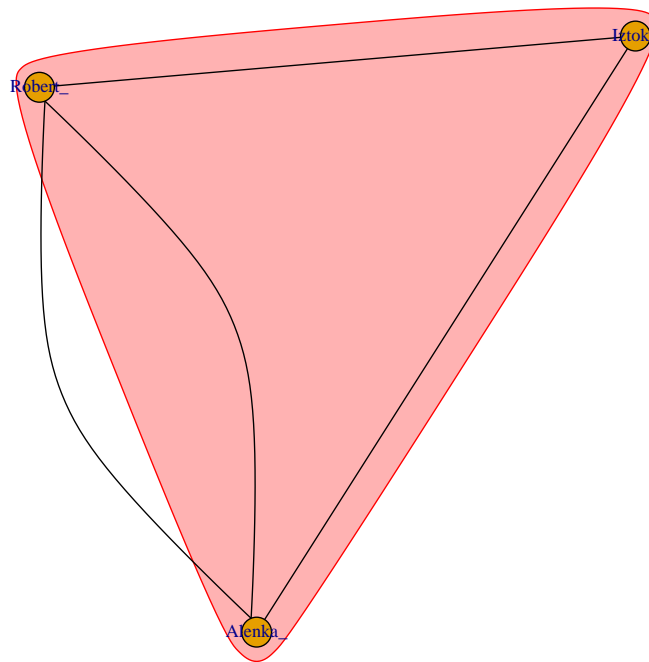
Programska koda zelo kratka. Na koncu je izrisan še graf povezav v socialnem omrežju, ki je shranjeno v bazi podatkov.

```
library(httr)
library(jsonlite)
library(igraph)
library(dplyr)
URL <- paste0("https://zgs6evrhemybfa-apexqxa8.",
              "adb.eu-frankfurt-1.oraclecloudapps.com",
              "/ords/fovcloud/mreza/vajaR")
rezultat <- GET(URL)
jsonRespParsed<-content(rezultat,as="parsed")
```

```

mreza <- as.data.frame(jsonRespParsed$items%>%
  bind_rows%>%
  select(vir,ponor,utez,vpis))
net <- graph.data.frame(subset(mreza,
  select = c(vir,ponor)),
  directed = F)
cnet <- cluster_edge_betweenness(net)
plot(cnet,
  net,
  vertex.size = 10,
  vertex.label.cex = 0.8)

```



4 Zaključek

V tej vaji so vključene kompleksne tehnologije kot so APEX, REST API in R. Prikazani primer kode v jeziku R ni zahteven, vendar pa je v ozadju tako aplikacija za zajem podatkov kot tudi storitev za pridobitev točno določenega dela shranjenih podatkov. Pravilne nastavitve vseh komponent omogočajo eno-

stavno uporabo. Če bi imeli več vhodnih podatkov, bi se na grafu povezav lepo prikazale gruče - podmnožice posameznikov, ki imajo več skupnih znanstev. Ena od knjižnic za analizo gruč se imenuje *cluster* in je uporabna pri problemih, kjer iščemo skupine elementov velike množice, med katerimi so razlike najmanjše. Praktični primer uporabe gručenja je npr. določanje skupine kupcev z zelo podobnimi lastnostmi, preferencami, nakupovalnimi navadami in podobno. Če koga zanima znanost o podatkih, je na voljo vrsto brezplačnih spletnih virov (npr. [2]) in veliko knjig (npr. [3]). Znanost o podatkih je interdisciplinarno področje, ki uporablja znanstvene metode, procese in algoritme za pridobitev znanja ali za odrivanje zakonitosti. Povezuje statistiko, podatkovno analizo in informatiko ter tudi podatkovno rudajenje, strojno učenje in masovne podatke.

Literatura

- [1] Dough Gault. REST and Application Express, May 17, 2019
- [2] rlogger. Novice in navodila za uporabo R
- [3] H. Wickham, G. Grolemund. R for Data Science