



Preddiplomski studij

Računarstvo

Informacija, logika i jezici

1. Laboratorijska vježba

Ak. g. 2023./2024.

Cilj ove laboratorijske vježbe je upoznavanje i samostalna izrada mrežne komunikacije klijent <-> poslužitelj. Dodatno, u vježbi će se obraditi parsiranje log datoteka te zadavanje upita nad njima. Za izradu upita koristit će se parser s odgovarajućom gramatikom.

Tokom izrade ove laboratorijske vježbe svi primjeri s predavanja se slobodno koriste i modificiraju po potrebi.

U programskom jeziku Java potrebno je napraviti klijenta i poslužitelja bazirane na protokolu TCP. Klijentski dio aplikacije potrebno je smjestiti u paket `hr.tel.fer.lab1.client`, a poslužiteljski dio u `hr.tel.fer.lab1.server`. Poslužitelj mora imati mogućnost rada s **više** klijenata. Maksimalni broj paralelnih klijenata, *port* na kojem će poslužitelj biti pokrenut i putanja do direktorija s log datotekama predaju se poslužitelju kao argumenti prilikom pokretanja. Možete pretpostaviti da se u danom direktoriju nalaze isključivo *.txt* datoteke sa zapisima jednakog formata. Nakon pokretanja, poslužitelj učitava log datoteke (koje je potrebno preuzeti sa stranica predmeta) te ih parsira i priprema za daljnju obradu. Poslužitelj nakon toga čeka spajanje klijenata i dobivanje upita.

Klase vezane za učitavanje i parsiranje log datoteke potrebno je smjestiti u paket `hr.tel.fer.lab1.logging`. Log datoteka sadrži podatke pristupanja poslužitelju www.hr te je strukturirana na sljedeći način: IPv4 adresa, datum i vrijeme, HTTP metoda, putanja kojoj se pristupalo, verzija HTTP protokola, statusni kod koji je vraćen klijentu te klijent koji je pristupio stranici. Separator između pojedinih elemenata je razmak. Pretpostavite da se zapis neće protezati kroz više redaka. IP adresa zapisana je u klasičnom formatu IPv4. Datum i vrijeme zapisani su u formatu [DD/MMM/YYYY:HH:mm:ss] npr. [12/Feb/2014:12:22:15]. HTTP metoda može biti jedna od sljedećih standardnih HTTP metoda: GET, PUT, POST, OPTIONS i HEAD. Putanja može biti navedena kao "*" ili započinje sa "/" nakon čega može i ne mora slijediti daljnja putanja (sljedeći zapisi su valjani: "/", "/putanja", "/putanja1/putanja2"). Verzija HTTP protokola zapisana je kao HTTP/1.0 ili HTTP/1.1. Statusni kod je troznamenkasti statusni kod HTTP protokola, npr. 200 ili 404. Klijent koji je pristupio stranici zapisan je unutar navodnika te može sadržavati razmake npr. "Apache/2.2.22 (Debian) (internal dummy connection)". Nakon dobivenog upita poslužitelj izvršava upit te korisniku vraća listu redaka koji zadovoljavaju uvjete upita. Za parsiranje upita obvezno je korištenje parsera s odgovarajućom gramatikom **uz pomoć alata ANTLR4**. Klase vezane za rad s upitima potrebno je smjestiti u programski paket `hr.tel.fer.lab1.query`.

Prilikom pokretanja klijenta kao argumenti predaju se IP adresa i vrata (*port*) te upit. Nakon slanja upita poslužitelju, klijent će kao odgovor dobiti listu redaka koji zadovoljavaju uvjet zadan upitom. Klijent ispisuje na zaslone dobiveni odgovor i završava s radom.

Upit koji se šalje mora biti zapisan u jednom retku. Upit je zapisan u određenom formatu i sadrži ključne riječi i operatore. Ključne riječi koje se pojavljuju su FILTER i RETURN koje se pojavljuju na početku, odnosno kraju upita te IP, DATETIME, METHOD, VERSION i STATUS koje se pojavljuju kod zadavanja uvjeta upita. Operatori koji se koriste u upitima su "==" i "!=". Opisi operatora i ključnih riječi navedeni su u sljedećoj tablici.

Ključna riječ/operator	Opis
FILTER	Svaki upit mora počinjati s riječi FILTER koja označava filtriranje po nekom uvjetu. Nakon riječi FILTER potrebno je navesti uvjet u formatu: <code>FILTER <ključna_riječ><operator><vrijednost></code> Moguće je nizati više uvjeta koji moraju biti odvojeni razmakom.
RETURN	Svaki upit završava s riječi RETURN koja označava broj zapisa koji će biti vraćeni te ima sljedeću sintaksu: <code>RETURN <value></code> Vrijednost koja se unosi mora biti ili prirodan broj ili znak * koji označava da je potrebno vratiti sve zapise koji odgovaraju uvjetu.
IP	Ključna riječ koja označava filtriranje po IP adresi. Moguće je uspoređivati po cijeloj IP adresi, ali također moguće je filtrirati po početnom dijelu adrese. Npr. ako se kao

	uvjet <code>IP=="161.53.XXX.XXX"</code> potrebno je vratiti sve retke čija adresa započinje s 161.53. Za ovu opciju dijelove IP adrese koji nedostaju potrebno je zapisati kao XXX.
DATETIME	Ključna riječ koja označava filtriranje po datumu i vremenu. Potrebno je uvijek zadati puni format kao vrijednost uvjeta, npr. <code>DATETIME=="12/Feb/2014:12:22:15"</code>
METHOD	Ključna riječ koja označava filtriranje po HTTP metodi. Kao vrijednost predaje se jedna od sljedećih metoda: GET, PUT, POST, OPTIONS, HEAD.
VERSION	Ključna riječ koja označava filtriranje po verziji HTTP protokola. Vrijednost koja se predaje u uvjet je u obliku <code>HTTP/<verzija></code>
STATUS	Ključna riječ koja označava filtriranje po statusu. Status je troznamenasti prirodni broj.
==	Operator koji označava jednakost uvjetu. Svi retci koji zadovoljavaju uvjet trebaju biti vraćeni korisniku, npr. <code>METHOD=="GET"</code> vraća sve retke gdje se koristi metoda GET.
!=	Operator koji označava različitost od zadanog uvjeta. Svi retci koji ne odgovaraju zadanom uvjetu trebaju biti vraćeni korisniku, npr. <code>METHOD!="GET"</code> vraća sve retke gdje se NE koristi metoda GET.

Ako format zadanog upita nije točan, potrebno je na adekvatan način obavijestiti korisnika.

Primjer nekoliko upita:

```
FILTER METHOD!="GET" STATUS=="200" RETURN"5"
```

Rezultat upita je prvih 5 (ili manje ukoliko ih nema 5) redaka koji kao HTTP metodu nisu koristili metodu GET, a dobiveni status je bio 200.

```
FILTER IP=="161.53.19.XXX" VERSION=="HTTP/1.1" RETURN"*"
```

Rezultat upita su svi redci koji zadovoljavaju uvjet da IP adresa počinje sa 161.53.19. te koriste verziju HTTP/1.1 protokola HTTP.

Navodnici navedeni u primjerima upita su opcionalni, kao i razmaci između ključnih riječi i operatora. Njihova uporaba u upitima mora biti konzistentna, tj. ne može ih se malo koristiti (npr. s ključnim riječima `METHOD` i `STATUS`), a onda ne koristiti s drugim ključnim riječima.