

Fakultet tehničkih nauka, DRA, Novi Sad

Predmet:
Organizacija podataka

Comma-Separated Values

CSV

CSV

- Comma-separated values
 - Datoteke zarezom odvojenih vrednosti
- Character-separated values
 - Datoteke znakom odvojenih vrednosti
- Tabelarni podaci predstavljani u obliku običnog teksta
 - Čist tekst bez formatiranja
 - Datoteke sa ekstenzijom *.txt ili *.csv
- Specifikacija data u [RFC-4180](#)

CSV

- Prva upotreba
 - 1967. u IBM kompajleru Fortran programskog jezika na operativnom sistemu OS/360
- Kasnija upotreba
 - Baze podataka, pregledači teksta
- Danas
 - Za razmenu podataka između baza podataka različitih arhitektura

CSV

- Prednosti
 - Čitljivost podataka putem tekstualnog editora

CSV

- Ne postoji standard za definisanje CSV datoteke, ali postoje pravila koja treba poštovati
 - CSV datoteka sadrži tabelarni zapis podataka
 - Svaki red u datoteci, definisan prelomom reda, predstavlja red u tabeli
 - Poslednji zapis na kraju može, ali ne mora, sadržati prelom reda
 - Prva linija može biti, ali ne mora, zaglavlje koje definiše naziv pojedine kolone

CSV

- U svakom redu datoteke se može nalaziti jedno ili više polja odvojeno znakom za odvajanje
 - svaki red treba da sadrži jednak broj polja
 - poslednje polje u redu ne sme biti popraćeno znakom za razdvajanje
- Razmaci su deo zapisa u polju i ne smeju se ignorisati
- Podaci u poljima reda mogu biti u dvostrukim navodnicima
- Polja koja sadrže podatke koji u sebi imaju znak za odvajanje, dvostruke navodnike i prelom reda moraju biti u dvostrukim navodnicima.
 - ako se dvostruki navodnici nalaze u podatku zapisanom u polju oni se moraju propratiti sa još jednim dvostrukim navodnicima

CSV

- Neki primeri odstupanja od pravila:
 - Razni znakovi se koriste za odvajanje npr. zarez (,), tačka zarez (;), tab (), razmak () i sl
 - Zapis praznog reda koji se može interpretirati kao zapis praznih i 0 vrednosti u svim poljima ili kao zapis u samo jednom polju
 - Korišćenje jednostrukih navodnika oko zapisa

Primer 1.

- Napisati *Java* program koji čita sadržaj datoteke *drzave_gradovi.csv* i ispisuje ga na ekran. Zadatak uraditi koristeći:
 - Klasu *java.util.Scanner*
 - Klasu *java.io.BufferedReader*
 - Klasu *CSVReader*

Konfiguracije prilikom čitanja

- Prilikom građenja objekta klase CSVReader, može se proslediti instanca klase CSVParser
 - Na taj način se može specificirati kako da se parsira pročitani sadržaj
 - specifikacija karaktera koji se koristi za razdvajanje vrednosti
 - specifikacija karaktera koji se koristi kao navodnik
- Takođe, može se specificirati koliko redova s početka bi trebalo preskočiti prilikom učitavanja podataka (preskakanje zaglavlja)

Primer 2.

- Napisati *Java* program koji formira izlaznu datoteku *drzave_gradovi2.csv* sa sledećim sadržajem:

Australia,Canberra

Canada,Ottawa

China,Beijing

France,Paris

Konfiguracije prilikom pisanja

- Prilikom kreiranja objekta klase CSVWriter, mogu se proslediti konfiguracioni parametri
 - Na taj način se može specificirati kako da se piše specificirani sadržaj
 - specifikacija karaktera koji se koristi za razdvajanje vrednosti,
 - specifikacija karaktera koji se koristi kao navodnik,
 - specifikacija karaktera kojim se izbegavaju specijalne vrednosti,
 - specifikacija karaktera kojim se završava red.

Primer 3.

- Pročitati sadržaj datoteke *drzave_gradovi.csv* u odgovarajuću listu *bean-ova* i ispisati podatke o učitanim gradovima.

Zadatak 1.

- Napisati (ručno) ulaznu csv datoteku, u kojoj svaki red datoteke sadrži x, y, z koordinate u prostoru za jednu tačku. Podaci su delimitirani znakom zarez.
 - a) Formirati csv datoteku koja u svakom redu sadrži x, y, z koordinate u prostoru za jednu tačku, kao i rastojanje te tačke od koordinatnog početka. Podaci su delimitirani znakom #.
 - b) Formirati izlaznu csv datoteku koja u svakom redu sadrži x, y, z koordinate u prostoru za jednu tačku, kao i rastojanje te tačke od koordinatnog početka. Tačke su sortirane u opadajućem redosledu prema rastojanju tačaka od koordinatnog početka. Podaci su delimitirani znakom #.

Zadatok 1.

- input.csv

x,y,z

1,2,3

3,4,5

5,6,7

....

output.csv

x#y#z#d

1#2#3#3.74..

3#4#5#7.07..

5#6#7#10.49..

....

Zadatak 2.

- Napraviti ulaznu csv datoteku, u kojoj svaki red datoteke sadrži broj indeksa, ime i prezime studenta, naziv predmeta koji je student polagao i ocenu koju student ima iz predmeta. Potom:
 - a) za svaku ocenu (5,6,7,8,9,10), formirati izlazne datoteke koje sadrže u svakom redu broj indeksa, ime i prezime studenta, i broj predmeta na kojima je student ostvario tu ocenu.
 - b) formirati izlaznu datoteku koja sadrži u svakom redu broj indeksa, ime i prezime studenta, broj predmeta koje je student položio, kao i prosečnu ocenu na položenim predmetima.

Zadatak 2.

br_indeksa	ime	prezime	predmet	ocena
1001	A	B	P1	6
1001	A	B	P2	7
1002	C	D	P1	8
1003	E	F	P3	10
1003	E	F	P2	7
1003	E	F	P1	7
1004	G	H	P1	5
1004	G	H	P2	6
1004	G	H	P3	7
1004	G	H	P4	8
1004	G	H	P5	9
1004	G	H	P6	10

b)

index	firstName	lastName	completedCou	avgScoreOnCom
1004	G	H	5	8
1003	E	F	3	8
1002	C	D	1	8
1001	A	B	2	6,5

a)

index	firstName	lastName	scoreCount
1004	G	H	1
1003	E	F	0
1002	C	D	0
1001	A	B	0

index	firstName	lastName	scoreCount
1004	G	H	1
1003	E	F	0
1002	C	D	0
1001	A	B	1

index	firstName	lastName	scoreCount
1004	G	H	1
1003	E	F	2
1002	C	D	0
1001	A	B	1

index	firstName	lastName	scoreCount
1004	G	H	1
1003	E	F	0
1002	C	D	1
1001	A	B	0

.

.

.

Zadatak 3.

- Data je ulazna csv datoteka (u direktorijumu dodatni_materijali u okviru materijala za termin), u kojoj svaki red datoteke sadrži broj indeksa, ime i prezime studenta, naziv predmeta koji je student polagao i ocenu koju student ima iz predmeta. Na početku datoteke nalazi se zaglavlje koje sadrži sledeće podatke:

Datum: Fakultet: Stud. Program:

Skolska godina: Semestar:

- Napisati *Java* program koji čita podatke iz datoteke i ispisuje kompletan sadržaj na ekran.

Zadatak 4.

- Napisati *Java* program putem kojeg će korisnik moći da opiše format csv datoteke: zaglavlje, broj kolona i nazivi kolona. Nakon toga omogućiti korisniku popunjavanje datoteke konkretnim podacima.