

# PRAKTIKUM IZ PROGRAMIRANJA 1

## - domaći zadatak broj 5 -

Sastaviti program na programskom jeziku Python, kojim se vrši određena vrsta obrade nad tekstualnim datotekama. Glavni program treba da:

- učitava nazive jedne ili više datoteka sa kojima će se raditi,
- poziva odgovarajuće funkcije za čitanje datoteke i obradu,
- ispisuje sve dobijene rezultate na način tražen postavkom zadatka;

Zavisno od rednog broja problema, svaki student treba da sastavi jedan od programa koji su dati u prilogu ovog dokumenta. Programi se bave obradom podataka o ATP teniskim mečevima u muškoj konkurenciji tokom 2021. godine.

Program treba da čita podatke uz vođenje računa o tipu podataka koji se čita. Svi programi čitaju tekstualnu datoteku u **Comma Separated Values (.CSV)** formatu koja sadrži podatke o teniskim mečevima u određenom periodu. Polja datoteke su razdvojena znakom zarez (,). Podaci se nalaze u datoteci čije se ime čita sa standardnog ulaza.

Datoteka ima zaglavlje sa nazivima polja koje treba ignorisati prilikom obrade. Podaci se u ulaznoj datoteci nalaze po sledećem formatu polja i značenju:

- Informacije o turniru
  - **tourney\_id** – identifikator turnira gde prva četiri karaktera označavaju godinu održavanja.
  - **tourney\_name** – naziv turnira.
  - **surface** – tip podloge koji može biti: *Hard* – tvrda, *Clay* – šljaka, *Carpet* – tepih ili *Grass* – trava.
  - **tourney\_date** – datum turnira u formatu *GGGGMMDD*, gde *GGGG* predstavlja godinu, *MM* mesec, a *DD* dan početka održavanja turnira.
  - **best\_of** – maksimalan broj setova na koji se meč igra.
- Informacije o igračima su date za pobjednika (prefiks **winner**) i gubitnika (prefiks **loser**)
  - **winner\_id/loser\_id** – jedinstveni identifikator igrača.
  - **winner\_name/loser\_name** – ime igrača.
  - **winner\_hand/loser\_hand** – dominantna ruka igrača. Može imati vrednost **R** za desnu ruku, **L** za levu ili **U** ako je informacija nepoznata.
  - **winner\_ht/loser\_ht** – visina igrača u centimetrima. Ukoliko podatak nije dostupan, polje je popunjeno vrednošću **N/A**.
  - **winner\_ioc/loser\_ioc** – troslovena oznaka države na osnovu kodiranja Međunarodnog olimpijskog komiteta koju igrač predstavlja.

- `winner_age/loser_age` – godine starosti igrača trenutku igranja turnira u obliku realnog broja.
- Informacije o meču
  - `score` – rezultat u setovima i gemovima. Podaci o setovima su dati u obliku `w-l`, gde `w` predstavlja broj gemova koje je pobednik, a `l` broj gemova koje je osvojio gubitnik. Ukoliko se skor završava sa `RET`, to označava da je gubitnik predao meč. Ukoliko umesto skora sotji `w/o`, to znači da meč nije odigran, a pobednik je automatski određen zbog povlačenja ili predaje gubitnika.
  - `minutes` – dužina meča u minutima.

Neka polja u ulaznoj datoteci mogu biti prazna. U tom slučaju tretirati ih kao ispravna i prilikom njihovog ispisa ispisati prazan string. Realne brojeve ispisivati na dve decimale.

Kompletan primer ulaza je dat u datoteci `atp_matches_2021.csv` u prilogu ovog dokumenta. Datoteka se može otvoriti za pregled u bilo kom editoru teksta ili pomoću *Microsoft Excel* alata. Za svaki problem su definisani i posebni, kraći ulazni primeri.

**Voditi računa o dekompoziciji programa na potprograme prema gore navedenoj raspodeli i zahtevima pojedinačnih zadataka.** Po potrebi, dozvoljeno je uvoditi i dodatne potprograme u rešenje. Student sam treba da definiše imena funkcija, kao i argumente potrebne za njihov rad. Voditi računa da se funkcijama dostave samo neophodni podaci. Funkcije ne smeju pristupati promenljivama glavnog programa direktno, već samo putem svojih argumenata.

**Prilikom rada programa proveravati i obraditi moguće greške korišćenjem mehanizma izuzetaka.** U slučaju greške pri radu sa datotekom, na standardnom izlazu ispisati poruku `DAT_GRESKA`. U slučaju greške prilikom konverzije podataka ili nekorektnog sadržaja nekog od polja (podaci o datumu meča, godinama starosti igrača ili dužini meča), na standardnom izlazu ispisati poruku `PODACI_GRESKA`. Zatim, korektno prekinuti izvršavanje programa. Proveru raditi samo za navedena polja.

Ako nešto u postavci zadatka nije dovoljno precizno definisano ili ako su neki od zahteva međusobno suprotstavljeni, usvojiti razumnu pretpostavku i rešiti zadatak korišćenjem te pretpostavke. Na samoj odbrani obavestiti demonstratora o usvojenoj pretpostavci ili pretpostavkama. Programski kod rešenja zadatka treba da bude uredno komentaran, tako da pri pregledu programa lako može biti uočeno šta radi bilo koja programska celina.

Radi boljeg testiranja programa, odaberi nekoliko dodatnih skupova podataka sa kojima će program biti testiran. Svaki primer treba da sadrži ulazne podatke i očekivani izlaz za te podatke.

### **Napomene:**

1. Rok za predaju petog domaćeg zadatka je **utorak, 11.01.2022.** putem kursa predmeta na **Moodle** platformi za elektronsko učenje. Tačan termin za predaju će biti naknadno definisan za sve studente. Termin će biti ograničenog vremenskog trajanja.
2. Domaći zadaci će biti testirani i ocenjivani korišćenjem javnih i tajnih testova, kao i demonstracijom funkcionalnosti na odbrani.
3. Studentima će nekoliko dana pre roka za predaju biti dostupno okruženje za testiranje rešenja domaćeg zadatka na Moodle platformi za elektronsko učenje korišćenjem javnih i tajnih testova.
4. Prilikom predaje domaćeg zadatka studenti će rešavati i kratak test znanja u vezi rešenja domaćeg zadatka i relevantnog gradiva iz oblasti programiranja na programskom jeziku Python koje obuhvata temu domaćeg zadatka.

5. Odbrana petog domaćeg zadatka će biti održana u **sredu, 12.01.2022.** Studenti mogu da se prijave za jedan od ponuđenih termina odbrane na Moodle platformi za elektronsko učenje.
6. Domaći zadaci se rešavaju **samostalno**. Predmetni nastavnici zadržavaju pravo da nakon predaje domaćih zadataka izvrše proveru sličnosti i preuzmu odgovarajuće disciplinske mere.
7. Svi drugi detalji oko predaje, ocenjivanja i odbrane domaćeg zadatka će biti blagovremeno objavljeni.
8. Formula za redni broj problema **i** koji treba rešavati je sledeća (R – redni broj indeksa, G – poslednje dve cifre godine upisa):

$$i = (R + G) \bmod 4$$

9. Kao rešenje domaćeg zadatka potrebno je predati sadržaj sledeće datoteke:
  - **dz5.py**, koja sadrži izvorni tekst programa na programskom jeziku Python;

31.12.2021. godine

*sa predmeta*

0. Napisati program koji vrši određenu obradu nad tekstualnom datotekom u **Comma Separated Values (.csv)** koja sadrži podatke o ATP teniskim mečevima u određenom periodu prema prethodno opisanom formatu. Program treba da formira izlaznu datoteku koja će za svakog tenisera koji je odigrao bar jedan meč odrediti najveći broj dobijenih mečeva na jednom turniru, kao i nazive takvih turnira, kao i ukupan broj odigranih mečeva u zadatom periodu. Jedan red izlazne datoteke treba da sadrži sledeća polja: string koji sadrži puno ime igrača (**player\_name**), najveći broj dobijenih mečeva na jednom turniru (**max\_wins**), listu sa nazivima turnira na kojima je ostvario najviše pobeda u zadatom vremenskom periodu uređenu leksikografski rastuće po nazivu turnira i razdvojenu znakom | (**top\_player\_tournament\_list**) i ukupan broj odigranih mečeva u zadatom periodu (**total\_matches\_played**). Izlazna datoteka treba da bude u **csv** formatu. U izlaznu datoteku na početku upisati zaglavlje. U izlaznu datoteku ne upisivati podatke o igračima koji nisu odigrali nijedan meč u zadatom periodu. Izlaznu datoteku urediti nerastuće prema ukupnom broju odigranih mečeva, a zatim leksikografski neopadajuće prema imenu igrača za one igrače sa istim brojem odigranih mečeva.

Program treba da:

- 1) Učita imena ulazne i izlazne datoteke sa standardnog ulaza.
- 2) Učita dva datuma u formatu **dd/mm** koji predstavljaju vremenski period za koji se vrši obrada. Svaki datum se unosi u posebnoj liniji teksta.
- 3) Izvrši zahtevanu obradu prema tekstu zadatka.
- 4) Formira izlaznu datoteku prema tekstu zadatka.
- 5) Vodi računa i obradi moguće izuzetke koji mogu nastati prilikom rada programa.

### Primeri

<b>Primer 1</b>	
<b>Standardni ulaz</b>	atp_few_2021.csv izlaz0_1.csv 01/01 01/07
<b>Ulazna datoteka:</b>	atp_few_2021.csv
<b>Izlazna datoteka:</b>	izlaz0_1.csv
<b>Standardni izlaz:</b>	
<b>Primer 2</b>	
<b>Standardni ulaz</b>	atp_selected_2021.csv izlaz0_2.csv 01/01 31/12
<b>Ulazna datoteka:</b>	atp_selected_2021.csv
<b>Izlazna datoteka:</b>	izlaz0_2.csv
<b>Standardni izlaz:</b>	
<b>Primer 3</b>	
<b>Standardni ulaz</b>	atp_matches_2021.csv izlaz0_3.csv 01/01 31/12
<b>Ulazna datoteka:</b>	atp_matches_2021.csv
<b>Izlazna datoteka:</b>	Izlaz0_3.csv
<b>Standardni izlaz:</b>	
<b>Primer 4</b>	
<b>Standardni ulaz</b>	nepostoji.csv izlaz.csv 01/01 01/03
<b>Ulazna datoteka:</b>	-
<b>Izlazna datoteka:</b>	-
<b>Standardni izlaz:</b>	DAT_GRESKA

1. Napisati program koji vrši određenu obradu nad tekstualnom datotekom u **Comma Separated Values (.csv)** koja sadrži podatke o ATP teniskim mečevima u određenom periodu prema prethodno opisanom formatu. Program treba da formira izlaznu datoteku koja će za zadatu podlogu turnira (**Hard, Clay, Carpet** ili **Grass**) i zadate godine starosti igrača (realan broj) odrediti sve pobednike turnira na zadatoj podlozi koji su mlađi od zadatih godina starosti u trenutku igranja turnira. Smatrati da je pobednik turnira onaj igrač koji je ostvario najviše pobeda na posmatranom turniru. Datoteka treba da bude uređena rastuće po starosti igrača u trenutku pobede na turniru. Jedan red izlazne datoteke treba da sadrži sledeća polja: naziv turnira (**tourney\_name**), ime igrača (**winner\_name**), starost igrača (**winner\_age**). Izlazna datoteka treba da bude u **csv** formatu. U izlaznu datoteku na početku upisati zaglavlje. Ukoliko na nekom turniru zadate podloge nije pobedio igrač mlađi od zadatih godina starosti, izostaviti ga iz izlazne datoteke. Podatak o starosti igrača zaokružiti na dve decimale.

Program treba da:

- 1) Učita imena ulazne i izlazne datoteke sa standardnog ulaza.
- 2) Učita podatke o podlozi turnira i maksimalnoj starosti igrača za koje se vrši obrada. Svaki podatak se unosi u posebnoj liniji teksta.
- 3) Izvrši zahtevanu obradu prema tekstu zadatka.
- 4) Formira izlaznu datoteku prema tekstu zadatka.
- 5) Vodi računa i obradi moguće izuzetke koji mogu nastati prilikom rada programa.

### Primeri

<b>Primer 1</b>	
<b>Standardni ulaz</b>	atp_few_2021.csv izlaz1_1.csv Hard 21
<b>Ulazna datoteka:</b>	atp_few_2021.csv
<b>Izlazna datoteka:</b>	izlaz1_1.csv
<b>Standardni izlaz:</b>	
<b>Primer 2</b>	
<b>Standardni ulaz</b>	atp_selected_2021.csv izlaz1_2.csv Clay 27
<b>Ulazna datoteka:</b>	atp_selected_2021.csv
<b>Izlazna datoteka:</b>	izlaz1_2.csv
<b>Standardni izlaz:</b>	
<b>Primer 3</b>	
<b>Standardni ulaz</b>	atp_matches_2021.csv izlaz1_3.csv Clay 35
<b>Ulazna datoteka:</b>	atp_matches_2021.csv
<b>Izlazna datoteka:</b>	izlaz1_3.csv
<b>Standardni izlaz:</b>	
<b>Primer 3</b>	
<b>Standardni ulaz</b>	nepostoji.csv izlaz.csv Clay 35
<b>Ulazna datoteka:</b>	-
<b>Izlazna datoteka:</b>	-
<b>Standardni izlaz:</b>	DAT_GRESKA

2. Napisati program koji vrši određenu obradu nad tekstualnom datotekom u **Comma Separated Values (.csv)** koja sadrži podatke o ATP teniskim mečevima u određenom periodu prema prethodno opisanom formatu. Program treba da formira izlaznu datoteku koja će na zadatoj podlozi (**Hard, Clay, Carpet** ili **Grass**), za svaku državu koja se pojavljuje u polju **winner\_ioc** ili **loser\_ioc** ulazne datoteke odrediti igrača sa najboljim odnosom broja pobeda u odnosu na ukupan broj odigranih mečeva na svim turnirima i igrača sa najlošijim odnosom broja pobeda u odnosu na ukupan broj odigranih mečeva. Podloga od interesa se unose iz jedne linije sa standardnog ulaza. Jedan red izlazne datoteke treba da sadrži sledeća polja: skraćeno ime države (**ioc**), ime igrača (**player\_name\_best**) sa najboljim odnosom broja pobeda u odnosu na ukupan broj odigranih mečeva (**player\_best\_w2p\_ratio**), ime igrača (**player\_name\_worst**) sa najlošijim odnosom broja pobeda u odnosu na ukupan broj odigranih mečeva (**player\_worst\_w2p\_ratio**). Ukoliko ima više takvih igrača, ispisati ime prvog u leksikografskom poretku. Izlazna datoteka treba da bude u **csv** formatu. U izlaznu datoteku na početku upisati zaglavlje. Poredak država u izlaznoj datoteci treba da bude uređen leksikografski rastuće po imenu države. Podatak o odnosu broja pobeda u odnosu na ukupan broj odigranih mečeva zaokružiti na dve decimale.

Program treba da:

- 1) Učita imena ulazne i izlazne datoteke sa standardnog ulaza.
- 2) Učita string koji sadrži podlogu za obradu.
- 3) Izvrši zahtevanu obradu prema tekstu zadatka.
- 4) Formira izlaznu datoteku prema tekstu zadatka.
- 5) Vodi računa i obradi moguće izuzetke koji mogu nastati prilikom rada programa.

### Primeri

<b>Primer 1</b>	
<b>Standardni ulaz</b>	atp_few_2021.csv izlaz2_1.csv Clay
<b>Ulazna datoteka:</b>	atp_few_2021.csv
<b>Izlazna datoteka:</b>	izlaz2_1.csv
<b>Standardni izlaz:</b>	
<b>Primer 2</b>	
<b>Standardni ulaz</b>	atp_selected_2021.csv izlaz2_2.csv Grass
<b>Ulazna datoteka:</b>	atp_selected_2021.csv
<b>Izlazna datoteka:</b>	izlaz2_2.csv
<b>Standardni izlaz:</b>	
<b>Primer 3</b>	
<b>Standardni ulaz</b>	atp_matches_2021.csv izlaz2_3.csv Hard
<b>Ulazna datoteka:</b>	atp_matches_2021.csv
<b>Izlazna datoteka:</b>	izlaz2_3.csv
<b>Standardni izlaz:</b>	
<b>Primer 4</b>	
<b>Standardni ulaz</b>	nepostoji.csv izlaz.csv Hard
<b>Ulazna datoteka:</b>	-
<b>Izlazna datoteka:</b>	-
<b>Standardni izlaz:</b>	DAT_GRESKA

3. Napisati program koji vrši određenu obradu nad tekstualnom datotekom u **Comma Separated Values (.csv)** koja sadrži podatke o ATP teniskim mečevima u određenom periodu prema prethodno opisanom formatu. Program treba da formira izlaznu datoteku koja će za svakog tenisera odrediti tri najduža meča koja je igrao u zadatom periodu. Jedan red izlazne datoteke treba da sadrži sledeća polja: string koji sadrži puno ime igrača (**player\_name**), listu sa dužinama tri najduža meča uređenu nerastuće po dužini meča i razdvojenu znakom | (**top\_3\_length**), listu sa imenima protivničkih igrača u najdužim mečevima razdvojenu znakom | u istom poretku kao u prethodnom polju (**top\_3\_length\_opponents**) i listu rezultata u obliku **RRR**, gde je **R** može biti **W**, ukoliko je odgovarajući meč dobio, a **L**, ukoliko je izgubio, u istom poretku kao u prethodnom polju (**top\_3\_length\_results**). Ukoliko ima više protivničkih igrača sa istom najdužom dužinom meča, ispisati ih po leksikografskom poretku, a najviše prva tri u listi. Izlazna datoteka treba da bude u **csv** formatu. U izlaznu datoteku na početku upisati zaglavlje. U izlaznu datoteku ne upisivati podatke o igračima koji nisu odigrali barem tri meča u zadatom periodu. Izlaznu datoteku urediti leksikografski rastuće prema imenu igrača.

Program treba da:

- 1) Učita imena ulazne i izlazne datoteke sa standardnog ulaza.
- 2) Učita dva datuma u formatu **d.m** koji predstavljaju vremenski period za koji se vrši obrada. Svaki datum se unosi u posebnoj liniji teksta.
- 3) Izvrši zahtevanu obradu prema tekstu zadatka.
- 4) Formira izlaznu datoteku prema tekstu zadatka.
- 5) Vodi računa i obradi moguće izuzetke koji mogu nastati prilikom rada programa.

### Primeri

<b>Primer 1</b>	
<b>Standardni ulaz</b>	atp_few_2021.csv izlaz3_1.csv 1.3 1.7
<b>Ulazna datoteka:</b>	atp_few_2021.csv
<b>Izlazna datoteka:</b>	izlaz3_1.csv
<b>Standardni izlaz:</b>	
<b>Primer 2</b>	
<b>Standardni ulaz</b>	atp_selected_2021.csv izlaz3_2.csv 1.3 15.9
<b>Ulazna datoteka:</b>	atp_selected_2021.csv
<b>Izlazna datoteka:</b>	izlaz3_2.csv
<b>Standardni izlaz:</b>	
<b>Primer 3</b>	
<b>Standardni ulaz</b>	atp_matches_2021.csv izlaz3_3.csv 1.1 31.12
<b>Ulazna datoteka:</b>	atp_matches_2021.csv
<b>Izlazna datoteka:</b>	izlaz3_3.csv
<b>Standardni izlaz:</b>	
<b>Primer 4</b>	
<b>Standardni ulaz</b>	nepostoji.csv izlaz.csv 1.1 10.12
<b>Ulazna datoteka:</b>	-
<b>Izlazna datoteka:</b>	-
<b>Standardni izlaz:</b>	DAT_GRESKA