

Četvrti veliki domaći zadatak

Napomene:

- 1) Odbrana: sreda, 09.01.2019. godine, 16:30 u laboratoriji 60.
- 2) Bira se prvi ili drugi problem.
- 3) Rešenje jednog od ova dva problema treba doneti na odbranu, u vidu fajla VDZ_4.py.
- 4) U zaglavlju fajla, pod komentarom, treba navesti ime, prezime i broj indeksa.
- 5) Rešenja treba da budu otporna na greške, odnosno program treba da detektuje i prijavi pogrešno unete vrednosti (van predviđenog opsega ili pogrešnog tipa).
- 6) Potrebno je osmisliti test primere i doneti ih na odbranu.

Prvi problem

Napisati Python skriptu koja pravi jedinstveni CSV fajl sa rezultatima na osnovu više ulaznih CSV fajlova. Svaki CSV fajl ima zaglavlje u kojem se nalaze nazivi kolona za taj fajl. U svakom od ulaznih fajlova postoji kolona sa brojem indeksa, koji mogu biti dati u različitim formatima: gg/bbbb, gg-bbbb, gggg/bbbb, gggg-bbbb i svaka od ovih mogućnost bez početnih nula za broj indeksa. Ime i prezime u fajlovima ne mora postojati, ali u slučaju da postoji, u svakom fajlu mora biti isto za isti broj indeksa. U svakom od ulaznih fajlova postoje poeni za jednu ili više obaveza na predmetu. Sve kolone koje ne sadrže indeks ili ime i prezime su kolone koje sadrže poene sa obaveza. Poeni sa jedne obaveze ne moraju da se nalaze u samo jednom fajlu, ali moraju biti isti u slučaju da postoje u više fajlova.

CSV: https://en.wikipedia.org/wiki/Comma-separated_values

Python i CSV: <https://docs.python.org/3.0/library/csv.html> (opciono korišćenje)

Drugi problem

Napisati Python skriptu/e za obradu i izvlačenje podataka o šahistima. U prilogu su dati fajlovi o standardnim rejtinzima o šahistima za decembar i novembar 2018. godine (fajlovi se mogu naći i na sajtu <https://ratings.fide.com/download.phtml>). Radi lakše obrade, potrebno je napisati nekoliko funkcija.

1. S obzirom na to da su fajlovi veliki, napisati funkciju "*filtriranje*" koja na osnovu argumenta *X* formira novu datoteku "*boljiOdX*", koja iz originalne datoteke na osnovu odgovarajuće kolone rejtinga (DEC18, NOV18) čuva samo one osobe sa rejtingom boljim od prosleđenog argumenta *X*. Da bi se radilo filtriranje, prethodno dobro proučiti format tekstualne datoteke koju treba parsirati. Parsiranje ne raditi regularnim izrazima (jednostavnije je bez). Pozvati funkciju filtriranje za *X*=1800. **[bonus 1 poen]** Napisati funkciju "*filtriranjeRegex*" koja koristi regularne izraze za obradu datoteke. Meriti vremena izvršavanja za obe funkcije (*filtriranje* i *filtriranjeRegex*) za istu vrednost argumenta *X* (npr. 1800), uporediti ih i zašto dolazi do razlike (ako dolazi do razlike). Proveriti da li su dobijene datoteke identične.
2. Napisati funkciju "*najboljihX*", koja na osnovu imena datoteke i prosleđenog prirodnog broja *X* kreira datoteku "*najboljihX*", koja sadrži samo *X* redova originalne datoteke, i to igrača koji imaju najveći rejting. Pozvati funkciju za datoteku dobijenu u prethodnom koraku ("*boljiOd1800*"), za *X* = 2000.
3. Napisati funkciju "*napraviRecnik*" koja od datoteke dobijene u koraku 2 ("*najboljih2000*") formira rečnik šahista, gde je ključ ID šahiste (kolona "ID number"), a vrednost koja se čuva treba u sebi da sadrži informacije o imenu igrača, federaciji kojoj pripada i rejtingu igrača (može imati i više informacija od ovoga).

4. Pomoću biblioteke *BeautifulSoup* napisati funkciju "obradiSahistu" kojoj se prosledi ID šahiste. Funkcija treba da formira link sa podacima o šahovskim mečevima koji je zadati šahista odigrao u zadatom mesecu, a zatim parsiranjem dobijene veb stranice ekstrahuje informacije o mečevima (na osnovu zahteva koji su dati u sledećoj tački).

Primer: Format linka ka stranici sa mečevima Fabijana Karuane (FIDE ID: 2020009) za novembar 2015: https://ratings.fide.com/individual_calculations.phtml?idnumber=2020009&rating_period=2015-11-01

Izgled veb stranice rezultujuće veb stranice, sa podacima o mečevima uokvirenim u crveno:

The screenshot shows the FIDE Chess Ratings website interface. The main content area displays three tables of individual calculations for November 2015. The first table is for the 'FIDE World Cup 2015' with a total change of -9.40. The second table is for the '2015 Millionaire Chess Open II' with a total change of -2.96. The third table is for the '31st European Chess Club Cup' with a total change of -2.96. Each table lists players with their ratings, performance, and changes. Red boxes highlight specific rows in each table, indicating the data points of interest.

5. Pozivom metode "obradiSahistu" za svaki od ključeva iz rečnika dobijenim u tački 3, dopuniti rečnik tako da on sadrži i za svakog šahistu spisak drugih šahista koji postoje u rečniku sa kojima ima odigrane mečeve u posmatranom mesecu (npr. decembar 2018), uz informaciju koliko imaju međusobno odigranih mečeva u tom periodu.
6. Obezbediti metode za nalaženje (ili ispis) najboljih N šahista po sledećim kriterijumima:
- Najviše protivnika u posmatranom mesecu
 - Najveći broj odigranih partija sa istim protivnikom u posmatranom mesecu
 - Najveći broj učešća na turnirima u posmatranom mesecu
 - Najveći prosečan rejting protivnika u posmatranom mesecu (uključiti i protivnike koji se možda ne nalaze u rečniku, tj. nisu među najboljih 2000 šahista)