

Praktikum iz programiranja 3

Drugi - kolokvijum

Vreme izrade: 105 minuta

Na **Desktop-u** u direktorijumu **Rad** kreirati direktorijum **ImePrezime_BrIndeksa_Godina** i unutar njega sačuvati programe koji sadrže rešenje datog zadatka. Rešenje se sastoji iz više programa (**singa.py**, **material.py**, **infotools.py**)

U radnom direktorijumu nalaze se folder **podaci** sa **JSON** i **XML** fajlovima koji čuvaju podatke o muzičkim numerama okačene na sajtu firme **Singa**. Kompanije želi da odradi različite statistike nad podacima, ali se javlja problem da su podaci predstavljeni u dva različita formata **JSON** i **XML**. Broj fajlova sa podacima unapred nije poznat, kao ni lokacija samih fajlova unutar foldera (mogu su pronaći na bilo kojoj dubini). Zbog velike količine podataka kompanija je posao podelila na dva nezavisna posla, koje će obraditi specijalizovani timovi unutar firme. Prva posao se odnosi na obradu **JSON** fajlova, dok se drugi odnosi na obradu **XML** fajlova. Za prvi posao je zadužen specijalizovan tim za obradu **JSON** fajlova pod nazivom **Material**, dok je za drugi posao zadužen specijalizovan tim za obradu **XML** fajlova po nazivom **Infotools**.

1. Zadatak [20 poena]

Kompanija zahteva da oba tima dostave izveštaj o umetnicima, njihovoj zbirnoj popularnosti (suma kolone **popularity**), kao i broju pesama u kojima dati autor učestvuje. Izveštaji moraju da budu u CSV formatu sa kolonama, **ImeAutora**, **ZbirnaPopularnost** i **BrojPesama** (**material-authors.csv** i **infotools-authors.csv**). Kompanija zna da jedan autor može da se pojavi u izveštaju tima **Material** i **Infotools**. Nakon dobijenih statistika kompanija (**singa.py**) sprovodi analizu podataka i određuje autora koji poseduje najveću srednju popularnost (ukoliko dva autora poseduju istu srednju vrednost, uzeti autora čije je ime manje po ASCII vrednosti).

2. Zadatak [20 poena]

Pored izveštaja o umetnicima, kompanije želi da dobije statistiku o broju pojavljivanja svakog žanra po godina (potrebno je odraditi statistiku samo za godine koje se pojavljuju u podacima.). Ukoliko se neki žanr ne pojavljuje ni u jednoj pesmi tokom neke godine, nije ga potrebno pomenuti u statistici. Krajnji rezultat ovih analiza je CSV fajlovi (**material-genres.csv** i **infotools-genres.csv**) u formatu **Godina**, **Zanr** i **BrojPesama**. Kompanija zna da ista kombinacija žanra i godine se može pojaviti u izveštaju tima **Material** i **Infotools**. Nakon dobijenih statistika kompanija (**singa.py**) sprovodi analizu podataka i za svaku godinu određuje žanr sa najviše pesama (ukoliko dva žanra poseduju isti broj pesama za jednu godinu, uzeti žanr čiji je naziv manje po ASCII vrednosti).

Napomena:

Zadatak mora biti rešen upotrebom **rečnika**. Prilikom izdvajanja imena autora koristiti funkciju `lower()` i `strip()` na sledeći način **imeZaposlenog.lower().strip()**. Sortiranje rečnika po ključu **sorted(recnik.items(), key=lambda x: x[0])**. **Ne koristiti apsolutne putanje. Program singa.py pokreće procese material.py i infotools.py i čeka njihov završetak.**

Primer statistike autora:

material- authors.csv		
ImeAutora	ZbirnaPopularnost	BrojPesama
Apulanta	300.00	3
Jay-Z	150.00	2
B Young	90.00	1

infotools- authors.csv		
ImeAutora	ZbirnaPopularnost	BrojPesama
Apulanta	90.00	1
Sam White	190.00	2

Izlaz (**singa.py**) za prvog zadatak za konkretne fajlove **material-authors.csv** i **infotools- authors.csv**:

Apulanta 97.50

Primer statistike žanrova:

material-genres.csv		
Godina	Zanr	BrojPesama
2019	Rock	3
2020	Rock	4
2020	Pop	5

infotools-genres.csv		
Godina	Zanr	BrojPesama
2019	Alternative	4
2020	Rock	3

Izlaz (**singa.py**) za drugi zadatak za konkretne fajlove **material-genres.csv** i **infotools-genres.csv**:

2019 Alternative 4

2020 Rock 7