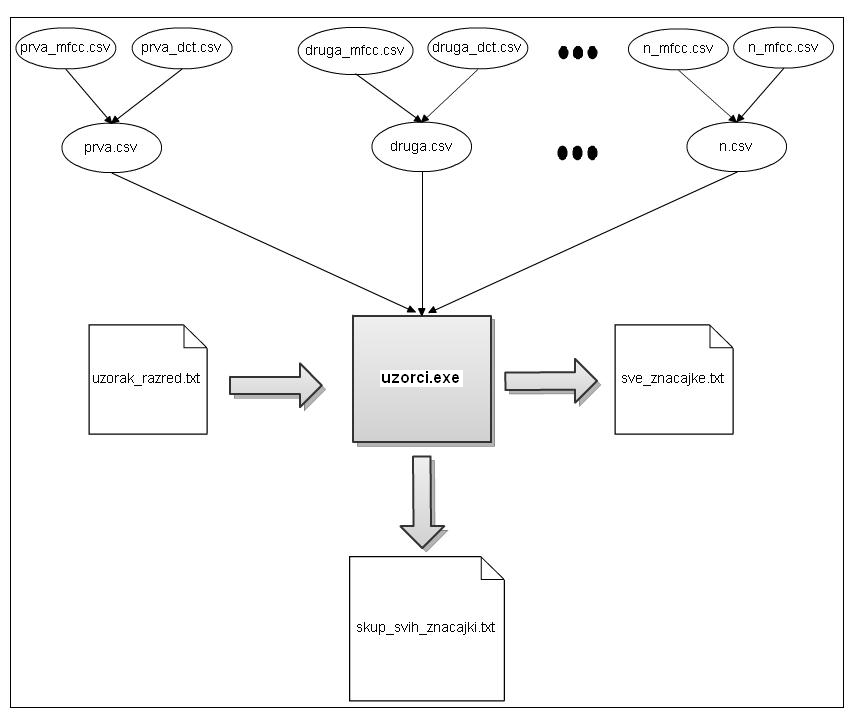
Nakon što smo izlučili značajke potrebno je tako dobivene vrijednosti zapisati na način prikladan za daljnju obradu. U ovom slučaju za treniranje i testiranje SVM-a.

Skripta „uzorci.pl“ radi upravo taj posao. Pronalazi sve datoteke u kojima su zapisane vrijednosti značajki (svaka pjesma ima svoju datoteku sa značajkama), te ih sprema u jednu datoteku. Za naš rad izabrali smo segmente pjesama u trajanju 5 sekundi. U prethodnom koraku moguće je odrediti koliko takvih segmenata želimo, upravo broj tih segmenata utječe na broj značajki. Uzmimo za primjer MFCC izlučivanje značajki, ono u segmentu od 5 sekundi sadrži 21 značajku, recimo da uzmemo tri proizvoljna segmenta od 5 sekundi tada ćemo imati 63 značajke.

Važno je spomenuti da navedena skripta ne ovisi o izboru značajki, te je moguće kombinirati više vrsta različitih značajki kako bi dobili što bolji rezultat. Pretraživanje različitih vrsta značajki je potpuno automatizirano i ukoliko postoji više vrsta značajki sve ćemo ih pripremiti za daljnji rad sa SVM-om.

Kako bi odredili kojem razredu pripada pojedini uzorak koristi se tekstualna datoteka koja sadrži imena svih pjesama s pripadajućom oznakom razreda.

Pogledajmo sada kako izgleda dio arhitekture sustava vezan za opisanu jedinicu posla:



Izvršni program "uzorci.exe" radi slijedeće:

1) uzima sve datoteke u trenutnom direktoriju i provjerava ako imaju ekstenziju .csv

2) zatim sve koji imaju tu ekstenziju spaja u jednu datoteku

3) prilikom stavljanja u jednu novu datoteku dodatno zapisuje podatak o kojem razredu

se radi, a to se provjerava u datoteci "uzorak\_razred.txt"

4) zatim sprema to sve u "sve\_znacajke.txt"

5) izabire značajke koje nam trebaju (one od 5 sekundi i pretvara ih u oblik

kojeg SVM može razumjeti) -> "skup\_svih\_znacajki.txt"

Upute za upotrebu:

Da bi program radio u istom direktoriju moraju biti:

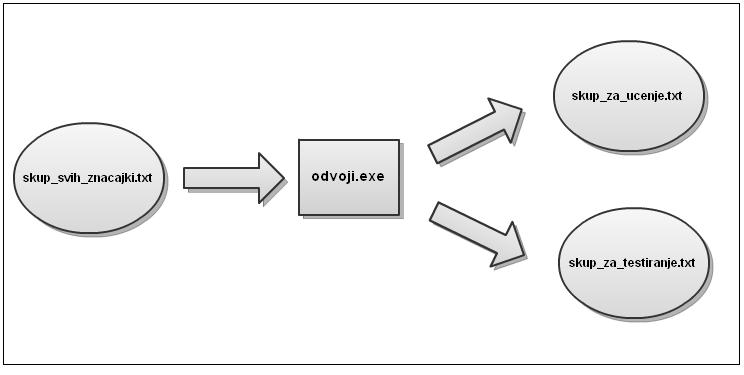
1) datoteke s ekstenzijom .csv

2) uzorci.exe

3) uzorak\_razred.txt

Nakon završetka rada u istom direktoriju stvorit će se datoteka s međurezultatima „sve\_znacajke.txt“ i datoteka s rezultatima koje koristimo za SVM „skup\_svih\_znacajki.txt“

Nakon što smo prethodno opisanim postupkom dobili datoteku sa pravilno oblikovanim značajkama još je potrebno razdvojiti uzorke za učenje i uzorke za testiranje. Ovaj postupak implementiran je u skripti odvoji.pl. Ova skripta kao argument prima postotak uzoraka koji će se koristiti za učenje. Ovisno o zadanom postotku uzoraka po razredu koji će se koristiti za učenje, datoteka „skup\_svih\_znacajki.txt“ će se razdvojiti na dvije nove datoteke. Jedna od njih je „skup\_za\_ucenje.txt“ u kojoj će biti zadani postotak uzoraka i druga novonastala datoteka je „skup\_za\_testiranje.txt“ u kojoj će biti svi preostali uzorci. Pogledajmo kako izgleda dio arhitekture koji se bavi ovim problemom:



Upute za upotrebu:

Poziva se iz ljuske slijedećom naredbom: odvoji.exe 50%

Ovaj program radi tako da iz "skup\_svih\_znacajki.txt" uzme prvih 50% uzoraka za svaki

razred, i spremi ih u "skup\_za\_ucenje.txt", a preostale spremi u "skup\_za\_testiranje.txt".

Primjer za parne brojeve: argument 50% -> ukupno 40 uzorka -> 20 za učenje,20 za testiranje

Primjer za neparne brojeve: argument 50% -> ukupno 33 uzorka -> 16 za učenje, 17 za testiranje

Dakle sve uzorke koji su ispod i jednako zadanom postotku stavi u učenje.

16/33 = 0,48 < 0,5 , dakle prvih 16 uzorka koristi se za učenje,

17/33 = 0,51 > 0,5 , a preostali uzorci se koriste za testiranje.