Projektni zadatak

Napraviti biblioteku za obradu podataka. Biblioteka treba da obezbedi efikasnu obradu podataka u CSV i JSON formatu. Pod obradom se podrazumeva primena operacija nad podacima u cilju dobijanja novih podataka. Sve operacije moraju za rezultat imati nove podatke i ne smeju dovoditi do izmene originalnih podataka. Operacije se mogu primenjivati proizvoljnim redosledom nad proizvoljnim podacima. Spisak osnovnih operacija koje je potrebno obezbediti dat je u tabeli 1.

Naziv operacije	Ulaz	Opis
filter	podaci, kriterijum	Formira nove podatke koji sadrže samo podskup originalnih podataka koji zadovljava zadati kriterijum. Kriterijum je proizvoljan.
transform	podaci, transformacija	Formira nove podatke koji su nastali primenom transformacije nad originalnim podacima. Transformacija je proizvoljna funkcija koja kao parametar prima jedan podatak iz prosleđenih podataka a za povratnu vrednost ima novu vrednost dobijenu primenom transformacije na prosleđeni podatak.
count	podaci	Vraća ukupan broj prosleđenih podataka.
countIf	podaci, kriterijum	Vraća ukupan broj podataka koji zadovoljavaju zadati kriterijum. Kriterijum je proizvoljan.
sum	podaci	Vraća sumu svih prosleđenih vrednosti.
average	podaci	Vraća srednju vrednost svih prosleđenih vrednosti.
order	podaci, ključ, *redosled, komparator*	Formira nove podatke koji su sortirani po zadatom ključu. Podrazumevani redosled je rastući. Ukoliko nije definisana komparator funkcija pretpostavlja se da se radi o poređenju nad primitivnim tipovima.
unique	podaci, ključ	Formira nove podatke u kojima se samo jednom pojavljuju originalni podaci čija vrednost ključa je jednaka.
join	podaci 1, podaci 2, ključ 1, *ključ 2*	Spaja podatke iz prvog skupa sa podacima iz drugog skupa po vrednostima zadatih ključeva. Rezultat su novi podaci sačinjeni iz oba skupa podataka. Ukoliko naziv drugog ključa nije naveden pretpostavlja se da je isti kao naziv prvog ključa.
append	podaci 1, podaci 2	Nadovezuje drugi skup podataka na prvi skup podataka.
upsert	podaci 1, podaci 2, ključ 1, *ključ 2*	Ukoliko vrednost ključa iz drugih podataka ne postoji u prvim podacima, dodaje podatke iz drugih podataka u prve, u suprotnom vrši izmenu vrednosti podataka u prvom. Rezultat je novi skup podataka sa izmenjenim i dodatim podacima. Ukoliko naziv drugog ključa nije naveden pretpostavlja se da je isti kao naziv prvog ključa.
split	podaci, kriterijum	Formira novi skup podataka u kojem su podaci podeljeni u dve grupe. Prva grupa sačinjena je od podataka koji zadovoljavaju kriterijum, dok je druga sačinjena od podataka koji ne zadovoljavaju kriterijum. Kriterijum je proizvoljan.
readCSV	putanja	Učitava podatke i vraća ih u formatu pogodnom za dalju obradu.
readJSON	putanja	Učitava podatke i vraća ih u formatu pogodnom za dalju obradu.
writeCSV	podaci, putanja	Zapisuje obrađene podatke u CSV formatu, i vraća zapisane podatke.
writeJSON	podaci, putanja	Zapisuje obrađene podatke u JSON formatu i vraća zapisane podatke.
plotSVG	podaci, putanja, handler	Na osnovu prosleđenih podataka formira grafik koji zapisuje u SVG formatu. Handler predstavlja funkciju koja vrši generisanje osnove SVG dokumenta i koja vraća funkciju za generisanje prikaza pojedinačnih elemenata iz prosleđenog skupa podataka.

Tabela 1 – Opis osnovnih operacija koje biblioteka mora podržavati. Nazivi ulaza napisani iskošenim tekstom predstavljaju neobavezne ulaze.

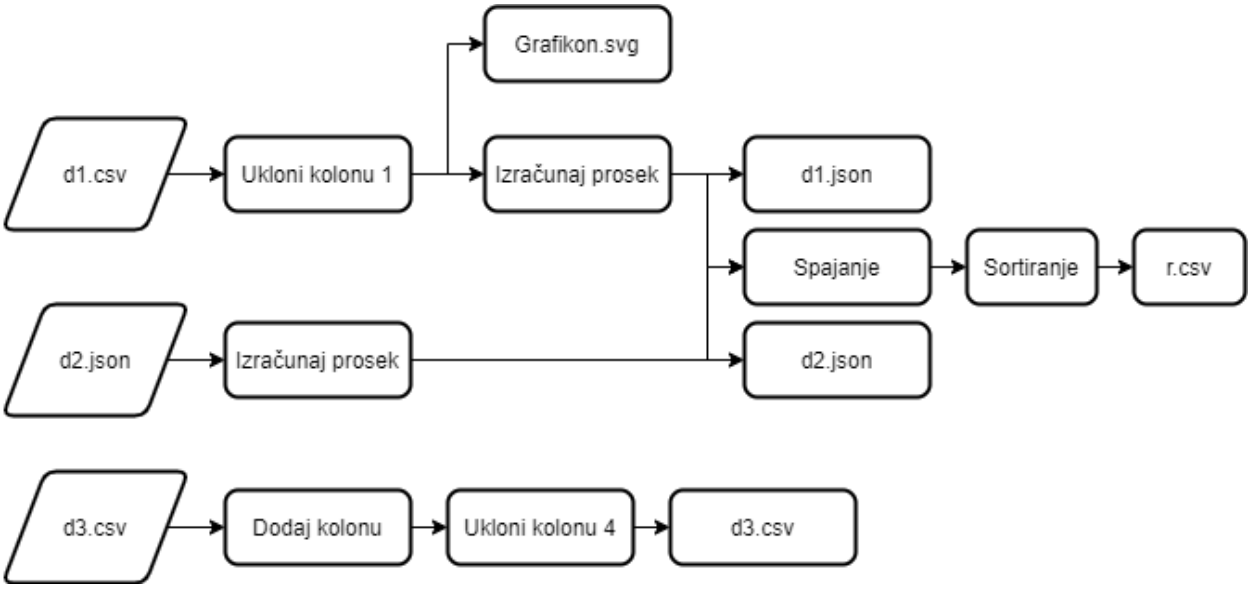
Biblioteka mora da podržava proširivost. Odnosno, neophodno je da se obezbedi mehanizam kojim bi se omogućilo kombinovanje postojećih operacija u složenije operacije i njihova ponovna upotreba. Na ovaj način bi trebalo da se omogući i proširivanje skupa tipova podataka koje je moguće čitati i pisati. Dodatno biblioteka treba da podržava i uslovno upravljanje tokom obrade podataka. Na tok obrade mogu uticati rezultati prethodnih obrada, npr. usled suviše malog broja preostlih podataka potrebno je izvršiti skup operacija 1 a ne skup operacija 2. Uslovi mogu biti proizvoljno zadati i opisuju se funkcijama koje dostavlja korisnik biblioteke.

Podaci kojima biblioteka manipuliše mogu biti sa različitih izvora i različitih veličina, takođe izvršavanje operacija nad različitim podacima može trajati različito vreme. Kako biblioteka treba da podržava obrađivanje više izvora podataka u isto vreme i spajanje rezultata ovih obrada potrebno je obezbediti mehanizam za asinhronu obradu podataka koji dozvoljava da operacije čekaju na pristizanje svih neophodnih podataka za njihovo izvršavanje. Takođe potrebno je omogućiti i paralenu obradu više podataka kroz upotrebu worker niti i sakupljanje rezultata više niti u jedan skup podataka.

Prilikom izvršavanja bilo koje od prethodno navedenih obrada podataka biblioteka treba da prati promene verzija podataka. Svaka nova verzija podatka treba da nosi podatak o verziji iz koje je proistekla, datumu i vremenu nastanka verzije, rednom broju verzije i operaciji koja je primenjena za dobijanje trenutne verzije. Kako biblioteka omogućava izvršavanje obrade više podataka u više niti potrebno je da se verzionisanje podataka vodi i na nivou pojedinačnih niti a ne samo celokupnog procesa.

Tokom obrade podataka može doći do neočekivanih otkaza usled neispravnosti podataka ili neispravnosti implementacija funkcija koje korisnik dostavlja biblioteci. Ovi otkazi ne smeju ugroziti celokupan proces obrade podataka. Proces bi bez obzira na otkaze trebao da nastavi izvršavanje do kraja i da u slučaju zapisivanja podataka na kojima je došlo do problema zapiše izveštaj o nastalim otkazima. Podaci na koje otkazi nisu imali uticaja treba da budu zapisani na uobičajen načni.

Slika 1 daje primer upotrebe biblioteke za obradu tri ulazna skupa podataka. Svaki skup podataka se obrađuje u odvojenoj niti pri čemu se rezultati obrade prvog i drugog skupa u jednom trenutku sakupljaju i dalje obrađuju kao jedna celina.



Slika 1 – Primer upotrebe biblioteke za obradu tri odvojena skupa podataka.