

SBNZ

Librarian

Sistem za preporuku knjiga za čitanje

Članovi tima:

- Strahinja Sekulić SV34/2020
- Lazar Milanović SV15/2020

Motivacija

Često nakon završene knjige ne znamo šta sledeće da čitamo. Ponekad želimo slične knjige, a nekad možda nešto malo drugačije. Uvek je teško odlučiti se za tu sledeću knjigu, jer ni u jednom slučaju nam niko ne garantuje da se u toj knjizi nalazi ono što baš nas zanima. Ili ćemo odabrati knjigu poznatog autora koja ne mora biti ni slična već pročitanim, ili biramo nasumično knjigu koja može, ali i ne mora, biti zanimljiva. Problem može da nastane i kod kupovine knjige za poklon jer često ne znamo šta bismo kupili osobi iako znamo šta voli da čita i šta je čitao/la ranije. Neophodan nam je dovoljno dobar način da odaberemo sledeću knjigu.

Pregled problema

Jedna opcija je otići u biblioteku gde bi nam stručna osoba preporučila knjigu, ali mnogi ljudi knjige kupuju preko interneta te im je ova opcija van dometa. Takođe, mnogo ljudi prosto ne želi da troši vreme na odlazak u biblioteku samo kako bi mu neko preporučio knjigu. Na internetu postoje i reader's advisory baze znanja koje preporučuju knjige, ali ili nisu dovoljno precizne, ili se plaćaju.

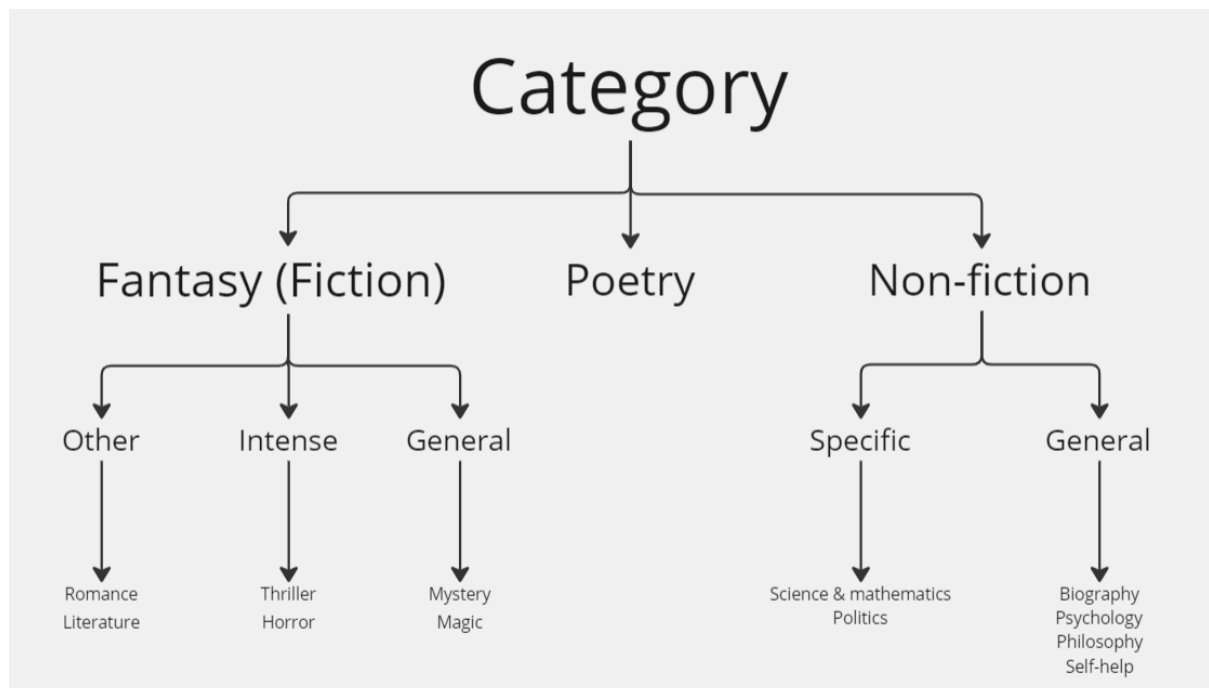
Naše rešenje ima za cilj da korisnicima olakša izbor sledeće knjige za čitanje. Sistem o korisniku prikuplja podatke. Od osnovnih podataka tu su pol i godište kako bi se vršile preporuke na osnovu slične demografije. Od korisnika se takođe zahteva da selektuje svoje preference - omiljenu kategoriju kao i neke od omiljenih ključnih pojmova. Korisnik takđe može da navede omiljene pisce kao i koja grupa godišta knjiga ga zanima kako bi dodatno suzio pretragu. Pored ovoga gledaju se i pročitane knjige kao i koji od pojmova iz njih je korisnik označio kao posebno dobre.

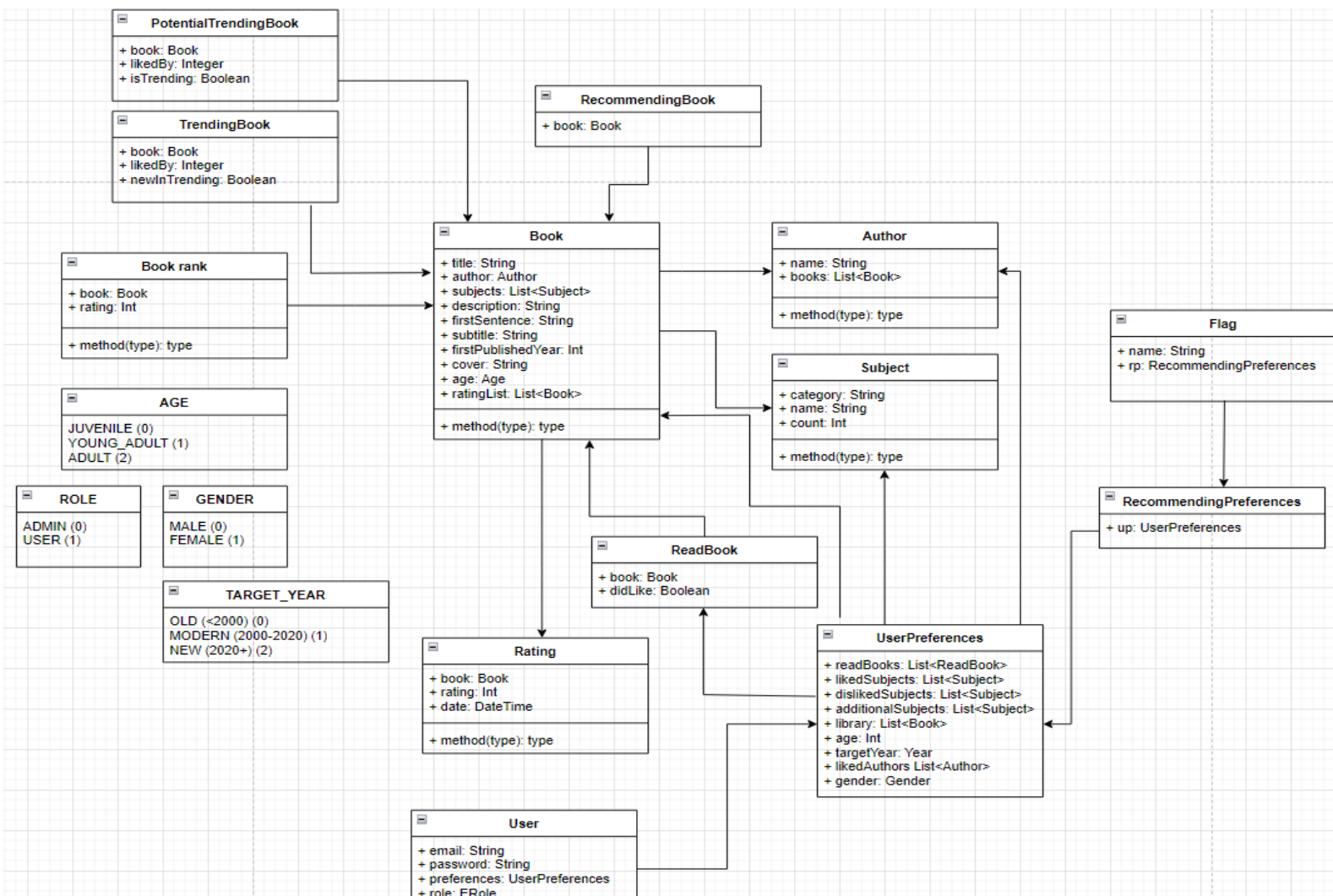
Unošenje većine podataka obavlja se pri registraciji korisnika te se kasnija preporuka knjiga svodi na prost izbor - da li korisnik želi sličnu knjigu već pročitanim, ili želi nešto drugačije. Sistem uzima sve činjenice, grubo filtrira knjige, ocenjuje ih i korisniku daje top 10 knjiga koje bi mogao sledeće da čita.

Sistem postaje sve precizniji što se više koristi jer on polako sve bolje uči šta korisnika zanima.

Metodologija rada

Svaka knjiga ima jednu ili više kategorija. Odabrali smo nekoliko "grubih" kategorija poput fikcije, dokumentarne literatura (non-fiction) i poezije. Te kategorije su dalje razložene na grupe kategorija. Konačno, na dnu stabla se nalaze precizne kategorije poput romantike, magije i biografije koje dosta blisko opisuju knjige. Uz ove kategorije, u našem skupu smo isfiltrirali oko 9000 ključnih pojmova koji se pominju u knjigama poput "truck drivers", "friendship", "mass media", "monsters", "zombie" i slično. Ključni pojmovi se sigurno nalaze u jednoj ili više pomenutih kategorija (imaju kategoriju kao parent-a). Na ovaj način ćemo omogućiti korisniku da preciznije bira pojmove koje bi želeo da čita, ili da u knjizi koju je pročitao obeleži pojam koji mu se najviše dopada. Takođe, na ovaj način smo u stanju da računamo sličnost između pojmova i tako preporučujemo knjige sa sličnim motivima.





- Book - Svaka knjiga ima naslov i autora. Poneke knjige imaju opis, prvu recenicu, sliku i podnaslov koji će nam pomoći u prikazivanju knjige na frontend-u aplikacije. firstPublishedYear je godina na osnovu koje ćemo primeniti pravila vezane za godine. Ako korisnik želi da čita najnovija izdanja ili starija izdanja to će moći da bira. Age je starosna grupa knjige. Knjiga može biti specifično za decu, za tinejdžere ili za sve uzraste. Uz knjigu je pridružena lista ključnih pojmova kao i lista ocena za tu knjigu.
- Author - Autor ima svoj naziv i listu knjiga koju je napisao. Taj podatak nam pomaže prilikom preporuke knjiga od istog autora.
- Subject - Ovde spadaju glavne kategorije i svi ključni pojmovi. Category predstavlja roditeljsku kategoriju dok naziv predstavlja naziv ključnog pojma. Uz subject je pridružen i broj ponavljanja koji predstavlja "relevantnost" motiva u odnosu na to koliko puta se taj motiv pojavljuje u ostalim knjigama u bazi podataka
- Book rank - Objekte ove klase će insertovati i menjati sam Drools gde će im dodavati relevantnost ili oduzimati relevantnost na osnovu pravila koje budemo definisali.
- Rating - Ovo je ocena knjige. Poneke knjige će imati ocenu čitaoca na osnovu kojih će im se povećavati relevantnost prilikom preporučivanja. Uvek ćemo prednost davati bolje ocenjenim knjigama

- User - Korisnik ima email i password na osnovu kojih se loguje na sistem. Po potrebi ćemo dodati i admina koji ima CRUD prava nad svim klasama aplikacije. Korisnik takođe ima i user preferences koje predstavljaju sviđanja tog korisnika.
- UserPreferences - Ovaj objekat nam služi da zaključimo koje knjige ćemo uzeti u obzir prilikom preporučivanja, a koje ćemo odbaciti.
- RecommendingPreferences - Ovaj objekat služi da se enkapsulira UserPreferences objekat za koji se vrši predviđanje
- RecommendingBook - Ovaj objekat služi da se enkapsulira Book objekat za koji se vrši predviđanje.
- Flag - Objekat koji koristimo u drools pravilima kako bismo znali koja je starosna grupa tog RecommendingPreferences-a
- PotentialTrendingBook - Pročitana knjiga sa brojem korisnika koji su je pročitali koja može postati TrendingBook.
- TrendingBook - Knjiga koja je već trending (pročitana je od strane najmanje 2 korisnika)

Baza znanja

Forward Chaining - preporuka knjige na osnovu preferenci korisnika:

1. Prvi nivo - Vršimo grubo filtriranje knjiga na osnovu uzrasta korisnika i glavne kategorije kako bismo smanjili početni skup koji posmatramo.
2. Drugi nivo - Svaku knjiga se modifikuje na osnovu preference godina izdavanja knjige, autora koji se korisniku sviđaju i kategorija koje mu se sviđaju. Dodajau se i dodatni poeni za dodatne pojmove koje želi da budu uključeni u buduću knjigu.
3. Treći nivo - Izbacuju se knjige koje su već pročitane i kažnjavaju se knjige koje imaju kategorije koje se korisniku ne sviđaju.
4. Četvrti nivo - Ovde modifikujemo knjige na osnovu njihovih ocena iz baze podataka koje su dali drugi korisnici, pružajući prednost knjigama sa boljom ocenom.
5. Peti nivo - Na kraju biramo 5 najbolje rangiranih knjiga i prikazujemo ih korisniku.

Forward Chaining - trending knjige

1. Prvi nivo - Kada je knjiga pročitana, sakupe se sve pročitane knjige i prebroji se koliko su puta pročitane.
2. Drugi nivo - Proverava se da li knjige ispunjavaju uslov za trending, što je da imaju najmanje 2 čitanja.
3. Treći nivo - Ako postoje knjige koje su prošle filter iznad, i ako postoje knjige koje su već trending, onda se sklanjaju knjige koje su već bile trending i postavljaju se nove knjige kako bi im se dala prednost.
4. Četvrti nivo - Na kraju se ažurira koliko je ljudi pročitalo knjige koje su već trending i čuva u bazi podataka.

Primer za backward

Preporuka knjiga na osnovu jedne pročitane knjige - sličnost po kategorijama

Za opciju koja pronalazi nekoliko sličnih knjiga na osnovu zadate tačno jedne knjige bez gledanja korisnikovih preferenci, korišćemo backwards chaining za pronalazak sličnosti po kategorijama. Kategorije su organizovane u stablo kategorija pa ćemo na osnovu toga da vidimo sličnost jedne kategorije sa drugim kategorijama. Akumulacijom ovih podataka i odabira top X najbližnjih kategorija, možemo preporučiti korisniku knjige koje sadrže baš te kategorije.

Izveštaj za najčitanije kategorije

Izveštaj je "o tebi" odnosno o korisniku. Tu se prikazuju njegove najčitanije kategorije na osnovu njegovih već pročitanih knjiga koje se dobivljaju korišćenjem backward chaining-a.

Primer za template

Kreiranje flagova za uzrast korisnika - kao ulazne parametre prima godine, naziv flag-a i salience. Kreiraju se ukupno 3 nova pravila.

Filtracija po glavnoj kategoriji - kao ulazne parametre prima samo kategoriju po kojoj se vrši filtriranje. Kreira se ukupno 18 novih pravila na ovaj način.

Filtracija po uzrastu korisnika - kao ulazne parametre naziv flag-a i EAge objekat. Kreiraju se ukupno 3 nova pravila.

Čišćenje flagova - kao ulazni parametar prima naziv flag-a. Ovde se kreiraju 3 nova pravila.

Čišćenje Recommending klasa - kao ulazni parametar prima naziv klase. Kreiraju se 2 nova pravila.