

SBNZ Predlog projekta

Book recommender system

Aleksandar Nikolić SW27/2017, Milan Jokanović SW28/2017

Motivacija

Ideja sistema je da olakša rad bibliotekama i poboljša user experience. Odlučiti se za narednu knjigu za čitanje može biti frustrirajuće, pogotovo kad je izbor praktično neograničen. Ideja ovog sistema je da reši taj problem i preporuči čitaocu narednu knjigu na osnovu njegovih preferencija. Pored preporuke čitaocima, sistem omogućava bibliotekama preporuku koji tip knjige je popularan i šta bi imalo smisla nabaviti u budućnosti.

Pregled problema

Ovaj sistem za cilj ima da pruži korisnicima biblioteka najbolje moguće iskustvo u potrazi za novom knjigom za čitanje. Takođe cilj sistema je da pomogne u administraciji biblioteke, tako što će davati izveštaj o najčitanijim knjigama, piscima i žanrovima preko kojih će admin sistema moći da zaključi koje nove knjige bi imalo smisla nabaviti.

Metodologija rada

- Ulaz
 - U slučaju preporuke za korisnika, ulaz je istorija tog korisnika, recenzije, knjige koje su drugi ljudi dobro ocenili a čitali su knjige kao i nadgledani korisnik itd.
 - Za izveštaj je ulaz generalna popularnost knjiga, pisaca, žanrova.
 - Na osnovu svakog poklapanja atributa knjige sa parametrom sa ulaza, povećava se ocena toj knjizi.
- Izlaz
 - Izlaz iz sistema na osnovu istorije čitanja korisnika je takođe 5 najbolje rangiranih knjiga.
 - Izlaz iz sistema za admina (biblioteku) su izveštaji najpopularnijih knjiga, pisaca i žanrova čiji bi cilj bio da pomognu pri odabiru knjiga za proširenje ponude.
- Baza znanja
 - Sistem će sadržati bazu dostupnih knjiga, koja će biti proširiva.
 - Svaka knjiga će sadržati attribute (žanrovi, autor(i), godina izdavanja, nobelovu nagradu, dužinu (broj stranica), da li je bazirana na pravom događaju, uzrast kojem je namenjena, ocene, broj pregleda, pripadnost serijalu)

- U sistemu će postojati korisnici i za svakog korisnika će se pamti koje je knjige pročitao, ocene dodeljene knjigama)
- Takođe će sistem u globalu pratiti najpopularnije knjige, pisce i žanrove i to znanje će se iskoristiti da ustanovi asistira u daljoj nabavci knjiga.

Pravila

Primer pravila 1 (automatska preporuka za ulogovanog korisnika):

Kada korisnik uđe na stranicu za automatske preporuke, okida se lanac pravila koji će vratiti 5 najbolje rangiranih knjiga koje korisnik nije pročitao na osnovu njegovih karakteristika.

Ocenjivanje po žanrovima

Povećanje skora svim knjigama koje sadrže žanr koji je korisnik u prošlosti čitao.

Formula: trenutni_skor + 0.5

Povećanje skora svim knjigama koje istovremeno sadrže više žanrova koje je korisnik u prošlosti čitao.

Formula: trenutni_skor + 0.1*broj_istovremenih_žanrova

Ocenjivanje po autorima

Povećanje ocene knjiga koje su napisali pisci koje je korisnik u prošlosti čitao na osnovu broj pročitanih knjiga od pisca i prosečne ocene tih knjiga (koje je on dao).

Formula:

$$\text{broj_pročitanih_knjiga_od_pisca} * (0.01 * \text{prosečna_korisnikova_ocena_pisca})$$

Ocenjivanje po sličnim korisnicima

Gledaju se ostali korisnici koji imaju iste pročitane knjige kao subjekt. Ostale knjige tih korisnika se skoruju na osnovu broja knjiga koje im se poklapaju. Kao i na osnovu proseka razlike ocena tih knjiga.

Primer proseka razlika:

KorisnikA dao ocene zajedničkim knjigama (5, 7)

KorisnikB deo ocene zajedničkim knjigama (8, 8)

Prosek razlika: $(\text{abs}(5-8) + \text{abs}(7-8)) / 2$

Formula:

$$(\text{broj_zajedničkih_knjiga} / 100) * (1 / \text{prosek_razlika_ocena_subjekta_korisnika})$$

Ocenjivanje nobelove nagrade i baziranosti na stvarnom događaju

Gleda se koliko je korisnik pročitao knjiga sa ovim osobinama i na osnovu toga se povećava skor onima koje poseduju ove osobine.

Formula:

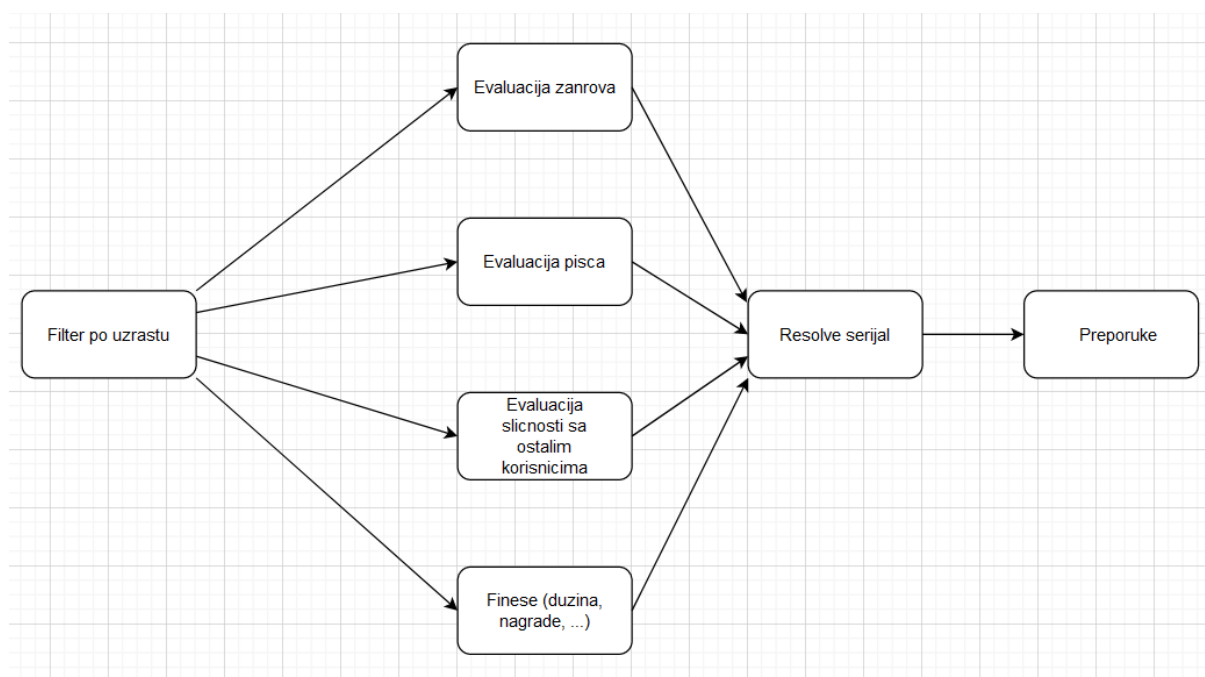
$$0.01 * \text{broj_pročitanih_knjiga_posmatranog_parametra (while } n < 10) + \\ 0.005 * \text{broj_pročitanih_knjiga_posmatranog_parametra (while } n > 10 \ \&\& \ n < 50) + \\ 0.001 * \text{broj_pročitanih_knjiga_posmatranog_parametra (while } n > 50)$$

Pronalaženje najboljih

Pronalazi 5 najbolje ocenjenih knjiga, s tim što pazi da se u listi nađe maksimalno 1 knjiga iz bilog koje serijala.

Handlovanje serijala

Ako se u listi najboljih našla knjiga serijala koja nije na redu da je korisnik čita (npr. Korisni je pročitao deo 1 i 2, a u listi se nađe deo 7), izbacije se i zameniti sa onom koja je korisniku na redu.



Slika 1. Tok izvršavanja pravila

Primer pravila 2 (izveštaj najpopularnijih knjiga u biblioteci):

Knjige delimo po njihovom broju stranica i potome da li su deo serijala ili ne. Kod serijala podele je samo da li jesu deo veceg serijala ili ne, dok podela po težini glasi.

Tipovi knjiga po dužini:

1. Kratke do 100 stranica
2. Srednje od 100 do 500 stranica

3. Dugačke od 500 stranica na više

Primer: Prilikom traženja najboljih 5 knjiga za izveštaj svaka knjiga će dobijati pojedinačnu ocenu i recimo da posmatramo "Lord of the Rings The Return of the King" koji ima 347 strana on bi spadao u knjige srednje dužine. U ovom slučaju nas interesuje kakve je zainteresovanost za čitanje knjiga srednje dužine, ovo dobijamo posmatranjem karakteristika svih knjiga u biblioteci. Takođe "Lord of the Rings" je veci serijala tako da bi za ovu knjigu računali i taj koeficijent.

Ulaz: Sve knjige iz sistema. Uzimamo redom knjige kako bi davali ocene.

1. Određujemo tipove knjiga na osnovu broja stranica, zatim kreiramo ocene za sva 3 tipa knjige kao sumu proizvoda prosečna ocene knjige i broja pregleda svih knjiga koje odgovaraju datom tipu, povećana vrednujemo knjige koje imaju preko pola svojih ocena većih od 6. Istu proceduru radimo kako bi smo generisali ocenu za knjige koje pripadaju serijalima, ali ovde sve knjige gledamo ravnopravno. Za serijale neće biti potreban korak za određivanje kategorija pošto se to nalazi u samom objektu knjige, računica ide po istoj formuli kao i za ocenu tipova dužine, ali ovde imamo samo 1 ocenu za to da li je knjiga deo serijala. Takođe na ovom nivou računamo generalnu prosečnu ocenu i prosečan broj pregleda svih knjiga.
2.
 - a. Sad kad svaki od tipova dužine ima ocenu dajemo im posebni koeficijent gde će dužina sa najboljom ocenom dobiti 0.5, sa srednjom 0 i ona sa najgorom ocenom -0.5, ako imamo dve dužine se istom ocenom obe dobijaju vrednost 0, postoji i opcija da sve 3 dužine dobiju koeficijent 0 ako su sve jednake. Slično kako imamo ocenu i za to da li je knjiga deo serijala ili ne i za to ćemo dati koeficijent gde ako knjige nije deo serijala on je 0, a ako jeste proveravamo da li je ocena za knjigu koje su deo serijala veći od proizvoda generalne prosečne ocene i prosečnog broja pregleda svih knjiga i ako jeste dajemo koeficijent 0.5, a ako nisu onda je -0.5.
 - b. Kako bi smo izračunali ocenu za knjigu takođe će nam biti potrebno da odredimo koeficijent za broj pregleda knjige. Ovo postizemo tako što ćemo posmatrati da li je broj pregleda veći od prosečnog broja pregleda koji smo prethodno izračunali. Iz ovoga dobijamo 3 pravila koji se dešavaju u sledećem redosledu. Ako je broj pregleda 2 puta veći od prosečnog broja pregleda onda je koeficijent 2, ako je samo veći, ali ne i 2 puta veći onda je 1 i ako je broj pregleda manji od prosečnog broja onda je 0.
 - c. Zelimo da dajemo prednost knjigama koje su favorit velikom broju čitalaca. Ovo postizemo ako knjiga ima broj favorit koji iznosi bar $\frac{1}{3}$ broja pregleda i ako je broj pregleda veći od prosečnog broja pregleda ili ako je broj favorita iznosi bar $\frac{2}{3}$ broja pregleda kad je broj pregleda manji od prosečnog, u ovim slučajevima koeficijent za favorite dobija

vrednost 1, dok ako nijedan od ova dva slučaja nije prošao onda je koeficijent 0, čitaoci koje brojimo su samo oni koji su bili aktivni u poslednjih 2 meseca.

3. Sad generišemo osnovnu ocenu knjige po formuli:

$$\text{ocena_knjige} * (\text{broj_recenzije} / \text{broj_pregleda}) + \text{koeficijent_favorita} + \text{koeficijent_broja_pregleda} + \text{koeficijent_tipa_dužine} + \text{koeficijent_serijal}$$

- 4.

- a. Sad kad imamo osnovnu ocenu knjige računamo ocenu žanrova na osnovu ocene knjiga koje imaju taj žanr po formuli:

$$\text{sum}(\text{osnovna_ocena_knjige} / \text{broj_žanrova_knjige})$$

Kad izracunamo ocene za sve žanrove takođe računamo koja je prosečna ocena svih žanrova pošto ćemo kasnije davati koeficijent koji žanr nosi na osnovu njih.

Sortiranjem žanrova po ovoj oceni dobijamo kakva je popularnost žanra.

- b. Za konačnu ocenu knjige želeli bismo da uključimo i ocenu za pisca, ali pre nego što ocenimo pisca prvo želimo da damo koeficijent za pisce koji su favoriti za bar 10% čitalaca u biblioteci. Koeficijent će imati vrednost 0 ili 1 u zavisnosti od toga da li je broj favorita veći od 10% ili ne, kao i kod knjige posmatramo samo čitaoce koji su bili aktivni u poslednjih 2 meseca kad brojimo favorite, ali 10% čitalaca se odnosi na sve čitaoce u sistemu čak i one koji nisu bili aktivni.

- c. Takođe da bi smo ocenili pisca moramo da nadujemo ukupnu ocenu za knjige koje su napisane od strane tog pisca ovo dobijamo po formuli
$$\text{suma}(\text{osnovna_ocena_knjige}) / \text{broj_napisanih_knjiga}$$
.

- d. Za kraj još dodeljujemo koeficije +1 ako smo ustanovili da pisac piše serijale gde knjiga pretežno dobijaju ocene čije su vrednosti preko 6 i da je broj takvih knjiga veći od $\frac{1}{3}$ svih knjiga koje je pisac napisao.

- e. Sad kad smo pripremili sve podatke dajemo ocenu i za pisce po formuli:

$$0.8 * \text{ocena_napisanih_knjiga} + \text{koeficijent_favorit_pisac} + \text{visoko_cenjeni_serijali_koeficijent}$$

- 5.

- a. Kako smo prethodno našli ocenu žanrova i iz toga mogli da pronađemo koja im je prosečna ocena, sad ćemo uz pomoć te 2 vrednosti da generišemo i koeficijent žanra za konkretne knjige. Pravilo glasi da ako je ocena žanra veća od prosečne onda se dodaje +1 u suprotnom ništa, ovo se ponavlja za sve žanrove koji se nalaze u knjizi.

6. Konačnu ocenu knjige se dobija po formuli:

$$0.8 * \text{osnovna_ocena_knjige} + 0.1 * \text{ocena_pisca} + (\text{koeficijent_žanrova_knjige}) / \text{broj_žanrova_knjige}$$

Konačna ocena knjige će se čuvati u okviru modela knjige kao sistemska ocena.

7. Sortiranje svih knjiga po sistemskoj oceni će nam dati 3 najbolje.

Računanje sistemskih ocena knjiga je generalno vremenski zahtevan zadatak, pogotovo ako imamo puno knjiga u sistemu. Kako čitaoci nebi morali da stalno čekaju dosta dugo da im se učitaju sistemski najbolje ocenjene knjige uvodimo CEP mehanizam kako bi se rešili ovog problema. Ideja je da se preko CEP-a proverava kad je bilo poslednje računanje sistemskih ocena i ako ovo vreme nije prešlo određeni prag dovoljno nam je samo da sortiranjem izvučemo knjige preko sistemskih ocene koje smo izračunali ranije. Tek kad ovo vreme pređe taj prag će biti potrebno da se ponovo računa sistemska ocena, jer smatramo da je moguće da je došlo do izmena koje mogu promeniti raspodelu knjiga u ovom vremenu, ali i ovo će se desiti samo prvom korisniku koji proba da preuzme sistemski knjige kad vreme istekne.

Još jedna bitna stvar za funkcionisanje našeg sistema jesu korisničke ocene koje su dodelili knjigama. Kao što se moglo i videti osnovni deo formule za računanje sistemske ocene jeste upravo prosečna ocena čitalaca. Uvidevši ovo jasno nam je da mora da postoji neki mehanizam koji će sprečiti čitaoce da unose ocene kad god hoće već da te ocene stvarno imaju neki značaj. Ovo smo takođe postigli korišćenjem CEP mehanizma, gde svako menjanje ocene od strane čitaoca kreira događaj ocenjivanja koji je vezan za njega. U trenutku kad broj ovih događaja predje prvi dozvoljen prag, sistem dodaje alarm događaj koji nam signalizira da čitalac ne postupa kako treba i da ga treba posmatrati. Ako čitalac nastavi sa politikom davanja ocene i ovo pređe drugi prag, a još uvek poseduje izdat alarm sposobnost čitaoca da daje ocene se blokira do daljnjeg. Događaji ocenjivanja i alarma će nestati posle nekog vremena, ali ako se korisnik blokira to neće nestati.

Literatura

Online Book Recommendation System using Collaborative Filtering (With Jaccard Similarity)- <https://iopscience.iop.org/article/10.1088/1742-6596/1362/1/012130/meta>