

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. Prednost u odnosu na dosadašnje radove jeste to što ćemo osim preporuka za knjige koje sistem daje na osnovu istorije čitanja korisnika omogućiti i opciju za custom pretragu knjige po parametrima. Ovo će omogućiti korisnicima da dođu do naslova koji nisu u skladu sa onim što su do sad čitali, a postalo im je od interesa. 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 mogao da nabavi.

Metodologija rada

- Ulaz
 - U slučaju preporuke preko forme ulaz su podaci sa same forme, sa tim što nisu sva polja u formi obavezna (tj. Korisnik može da ignoriše neki parametar tako što ga neće uneti). Forma će nuditi unos željenih žanrova, pisaca, opseg godina izdavanja, dužinu, uzrast kojem je namenjena, baziranost na pravom događaju, minimalnu prosečnu ocenu, nagrade, broj recenzija, pripadnost serijalu i književnu epohu.
 - 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 nakon traženja preporuke putem forme je 5 najbolje rangiranih knjiga.

- 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, dobijene nagrade, dužinu (broj stranica), period pripadanja, da li je bazirana na pravom događaju, uzrast kojem je namenjena, prosečnu ocenu, broj ocena, pripadnost serijalu)
 - U sistemu će postojati korisnici i za svakog korisnika će se pamtit i koje je knjige pročitao, ocene dodeljene knjigama, najčitaniji pisci i žanrovi, da li voli ili izbegava serijale)
 - 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. Na prvom nivou rezonovanja sistem filtrira knjige na osnovu starosne grupe čitaoca. Na drugom nivou rezonovanja sistem ocenjuje knjige na osnovu žanrova koje je korisnik čitao. Potom ocenjuje pisce koje je čitao kao i pisce koje nije čitao, a koji pripadaju žanru od interesovanja. Na četvrtom nivou gleda čitaoca (sličnog uzrasta) koji imaju zajedničke knjige sa subjektom (bonus ako su i subjekat i drugi čitalac dali visoku ocenu knjizi) i ocenjuje njihove ostale knjige (koje subjekat nije pročitao). Na petom nivou rezonovanja, traži trend u istoriji čitaoca za podatke o dužini knjige, nagradama koje je knjiga osvojila, da li je bazirana na istinitom događaju, pripadnost određenom književnom stilu, postojanje ekranizacije. Parametri na petom nivou su manje bitni i služe za finu korekciju preporuka. Na poslednjem nivou rezonovanja sistem prepoznaje da li je neka knjiga, u 5 najbolje rangiranih, članica nekog serijala, i ako jeste, automatski se zamenjuje sa prvom knjigom tog serijala koju korisnik nije čitao (kako bi korisnik uvek čitao serijal u odgovarajućem redosledu). Čitalac ima mogućnost da kaže sistemu da ignoriše neke od ovih parametara.

Primer: Recimo da je čitalac star 13 godina. U prvom koraku se filtriraju knjige za ovu starosnu grupu. Recimo da je korisnik čitao i visoko ocenio serijale "Percy Jackson" (fantastika, grčka mitologija, akcija, avantura), "Narnia" (fantastika, avantura) i "The tapestry series" (fantastika, avantura, school, akcija). Uočavamo da svi ovi serijali imaju žanr fantastike i avanture, te ostalim knjigama sa tim žanrovima povećavamo ocenu. Potom gledamo pisce i povećavamo ocenu njihovim ostalim knjigama, težina ovog koraka je srazmerna broju pročitanih knjiga određenog pisca.

Potom gledamo druge korisnike i recimo da imamo korisnike koji su pročitali i visoko ocenili sve ovo plus "Harry Potter" serijal i "Alex rider" serijal. Ovde vidimo da serijal "Harry Potter" ima jako poklapanje po žanrovima, i vidimo poklapanje ocena sa serijalima koje je i subjekat pročitao. Nakon ovog nivoa gledaju se finesse do tad rangiranih knjiga (dužina, nagrade, ...). Finalno se vrši filter za serijal i korisnik dobija listu 5 najbolje ocenjenih knjiga za njega u tom trenutku. To bi u ovom slučaju bio recimo "Harry Potter" serijal, gde bi sistem preporučio prvu knjigu serijala koju subjekat nije pročitao, iako je neka kasnija potencijalno imala bolju ocenu.

Primer pravila 2 (izveštaj najpopularnijih knjiga u biblioteci):

Kada admin zatraži izveštaj za najpopularnije knjige okida se lanac pravila koji vraća 5 najpopularnije rangiranih knjiga na nivou cele biblioteke. Na prvom nivou rezonovanja određujemo pripadnost knjige po dužini na osnovu broja stranica. Na drugom nivou rezonovanja sistem ocenjuje zajedničku vrednost za srednju ocenu knjiga i broj recenzija izdatih za datu knjigu, kao i broj pregleda knjige, broj čitalaca koji ima knjigu kao omiljenu. Na ovom nivou rezonovanja takođe dajemo ocenu za tipove dužina knjiga računanjem ukupnih pregleda knjiga tog tipa i njihovih prosečnih ocena, takođe ćemo ocenjivati to kako čitaoci vide knjige koje se deo većeg serijala. Na trećem nivou rezonovanja vršimo spajanja svih rezultata sa drugog nivoa u jednu ocenu koja će predstavljati izvedenu ocenu knjige. Na četvrtom nivou recenzije određujemo ocenu žanra na osnovu ocena knjiga sa trećeg nivoa rezonovanja, pri tom veću će ocenu dobijati od knjiga koje imaju manje žanrova, na ovom nivou ćemo takođe izdvojiti žanrove koje su pisci najčešće pisali i za koliko čitaoca je pisac omiljen. Na petom nivou rezonovanja dajemo ocenu piscu na osnovu ocena knjiga koje je napisao, broja čitalaca za koje je on omiljen i ocene žanrova koje je najčešće pisao. Na šestom nivou rezonovanja dobijamo konačnu ocenu za knjigu spojem izvedene ocene knjige sa ocenom žanrova i ocenom za pisca. Na sedmom nivou rezonovanja ćemo na osnovu dobijen konačne ocene proslediti adminu top 5 knjiga. 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. Formula za ocenjivanje je da ako ima mali broj pregleda i loša je prosečna ocena knjige onda umanjujemo ocenu za knjige srednje dužine, ako je bar jedan od prosečne ocene ili broj pregleda zadovoljavajući, a drugi loš onda se ta knjiga neće uzimati u obzir i ako su oba parametra zadovoljavajuća ocene će se povećati. Za "Lord of the Rings" takođe moramo da proverimo kakva je zainteresovanost čitalaca za knjige koje su deo serijala. Formula za ovo je ista kao i kad smo ocenjivali tipove dužine knjige, ali za

razliku od toga što je svaki tip dužine dobijao svoju ocenu, kad proveravamo samo knjige koje su deo serijala, ako knjiga nije deo serijala ovo pravilo neće uticati na njenu konačnu ocenu. Najbitniji deo ocene knjige jeste ocena koju dobijamo računanjem ukupnog broja pregleda knjige zajedno sa prosečnom ocenom i brojem čitalaca koji su dali ocenu. To znači da iako je knjiga imala dosta pregleda ako joj je prosečna ocena loša njena konačna ocena će opasti, ali ovo će jako zavisići od toga koliko je čitalaca stvarno ocenilo knjigu, jel ako je mali broj čitalaca ocenilo knjigu loše, a broj pregleda je veliki to ne znači da je ona stvarno loša samo da većina čitalaca nije imao nikakvo posebno mišljenje, ovo će umanjiti ocenu knjige, ali ne i u toj meri u kojoj bi bilo ako je broj korisnika koji je ocenio knjigu kao loša bio veći. U manjoj meri, ali definitivno još uvek važno je činjenica da li je knjiga smatra kao omiljena od velikog broja čitalaca. Knjiga može imati mali broj pregleda, ali ako je kod velikog broja tih pregleda smatra tu knjigu kao omiljenu za čitanje to znači da veliki broj čitalaca nije pronašao i da bi je trebalo preporučiti. Skupljanjem svih ocena dobićemo izvedenu ocenu knjige koju dalje koristimo kako bi ocenili žanrove i pisce vezane za te knjige. Za konačnu ocenu knjiga najbitnija nam je izvedena ocena za knjigu, zatim njen pisac i žanrovi, pri tome što je knjiga sadržana od više žanrova to je njihov pojedinačni uticaj manji.

Literatura

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