

Preporuka filmova korišćenjem collaborative filtering-a

-specifikacija projekta-

RA22-2013 Svetozar Stojković

Potrebno:

- Instaliran python 2.7 na računaru
- Biblioteke:
 - Python imdb api - `sudo apt-get install python-imdbpy`
 - Flask - `pip install flask`

Pokretanje:

- Pokrenuti Recommender file

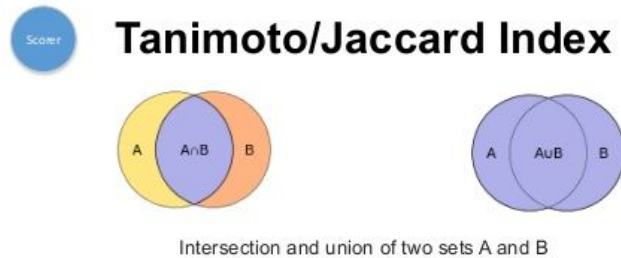


Funkcionalnosti:

- Registracija novog korisnika na sistem
 - Korisnik unosi username, password i ime čime se on doda u matricu korisnika i filmova.
- Logovanje korisnika na sistem
 - Korisnik se loguje sa adekvatnim kredencijalima na sistem i pristupa njegovom profilu gde se nalaze filmovi koje je korisnik ocenio.
- Pretraživanje svih filmova
 - Korisniku je prikazana tabela sa svim filmovima koji se nalaze u sistemu (33040 filmova u ovom projektu) koje on može da filtrira i bira koji želi.
- Ocenjivanje izabranog filma
 - Nakon što izabere odgovarajući film pojavi mu se tabela u kojoj se nalaze osnovni podaci o filmu: naziv, godina, slika, radnja i žanrovi. U toj tabeli se takođe nalazi opcija da korisnik oceni film.
- Dobavljanje preporuka izabranim algoritmom
 - Korisnik ima opciju da izabere algoritam koji će da generiše listu filmova preporučenih na osnovu njegovih ocenjivanja.

Algoritmi za preporučivanje:

- Žakar (Jaccard)
 - Jaccard algoritam računa sličnost između korisnika u ovom slučaju tako što računa presek interesovanja između korisnika podeljen sa unijom ocenjenih filmova.



$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$



https://en.wikipedia.org/wiki/Jaccard_index

- Kosinusni (Cosine)
 - Kosinusni algoritam funkcioniše tako što određuje sličnost između dva vektora na osnovu vrednosti kosinusa

```
similarity = 1 - spatial.distance.cosine(my_movies, other_user_movies)
```

- Centrirani kosinusni (Centered cosine)
 - Centrirani kosinus radi tako što od svih ne-nula vrednosti oduzima prosek ne-nula vrednost u vektoru odnosno nizu.
- Pearson:

$$r_{au} = \frac{\sum_{i=1}^n (r_{ai} - \bar{r}_a)(r_{ui} - \bar{r}_u)}{\sqrt{\sum_{i=1}^n (r_{ai} - \bar{r}_a)^2} \sqrt{\sum_{i=1}^n (r_{ui} - \bar{r}_u)^2}}$$

- r_{au} - koeficijent korelacije korisnika a i u
- r_{ai} - rejting stavke i korisnika a
- \bar{r}_a - srednja vrednost rejtinga