



**Diplomski studij**

**Informacijska i komunikacijska  
tehnologija**

Telekomunikacije i informatika

**Računarstvo**

Programsko inženjerstvo i  
informacijski sustavi

Računalno inženjerstvo  
Računarska znanost

## **Raspodijeljena obrada velike količine podataka**

### **3. Laboratorijska vježba**

**Ak. g. 2017./2018.**

## Zadatak 1: Implementacija sučelja ItemSimilarity koja sadrži sličnost objekata temeljenih na ocjenama korisnika

Za potrebe laboratorijskih vježbi koristit ćete datoteke kao i u zadaći (`jester_ratings.dat`) da biste izračunali sličnost objekata (šala) na temelju ocjena korisnika.

### Zadatak

Cilj zadatka je napisati implementaciju sučelja `ItemSimilarity` koje sadrži podatke o sličnosti šala na temelju korisničkih ocjena.

Uspješnim rješenjem ovog zadatka steći ćete sljedeća znanja:

- **izrada matrice sličnosti na temelju korisničkih ocjena**
- **korištenje dijela programskog okvira Mahout za izračun sličnosti**

### Detaljni opis zadatka

Proučite i proširite dobivenu implementaciju sučelja `ItemSimilarity` (`CollaborativeItemSimilarity.java`), tj. implementirajte sljedeće tri metode: `itemSimilarity`, `itemSimilarities` i `allSimilarItemIDs`. Opis navedenih metoda možete pronaći na [poveznici](#).

Klasa `CollaborativeItemSimilarity` sadrži sljedeće strukture podataka:

- `double[][] matrix` – matrica sličnosti šala
- `Map<Integer, Long> seqIdMap` – mapa koja preslikava sekvencijski broj šale na njen identifikator
- `Map<Long, Integer> idSeqMap` – mapa koja preslikava identifikator šale u njen sekvencijski broj.

Sekvencijski brojevi se koriste kao identifikatori šale samo unutar klase `CollaborativeItemSimilarity.java`.

## Zadatak 2: Izrada hibridne matrice sličnosti

(*NAPOMENA: kao ulaznu datoteku u zadatku koristite izlaznu strukturu podataka iz prethodnog zadatka i datoteku sličnosti šala temeljem sadržaja koju ste kreirali u domaćoj zadaći*)

### Zadatak

Cilj zadatka je stvoriti strukturu podataka koja će obuhvaćati sličnost šala temeljem ocjena korisnika i sadržaja. U tu svrhu maksimalno iskoristite prethodno kreirane strukture.

Uspješnim rješenjem ovog zadatka steći ćete sljedeća znanja:

- **izrada hibridne matrice sličnosti šala**

### Detaljni opis zadatka

U ovom zadatku potrebno je napraviti jedinstvenu matricu sličnosti koja obuhvaća sličnost temeljem sadržaja šala i sličnost temeljem ocjena korisnika. Takva hibridna matrica sličnosti dobiva se kombiniranjem dviju jednostavnih matrica koristeći definirane težinske faktore.

Izrada hibridne matrice se sastoji od sljedećih koraka

1. izrada matrica sličnosti temeljem jednog kriterija (već napravljeno)
2. normalizacija pojedinih matrica sličnosti u raspon `[0,1]`

### 3. zbrajanje pojedinih matrica sličnosti uz korištenje težinskog faktora

Prvi korak ste napravili kroz domaću zadaću i prvi zadatak. Od sada nadalje smatramo da imate pripremljene strukture `ItemSimilarity` temeljene na sadržaju i ocjenama korisnika.

U drugom koraku potrebno je normalizirati pojedine matrice sličnosti tako da budu usporedive. Svaka pojedina matrica se normira zasebno. Normalizacija će nam omogućiti da sami odredimo težinski faktor udjela pojedine sličnosti u hibridnoj matrici. Konačni rezultat normiranja će stvoriti matrice u kojima zbroj sličnosti jedne šale sa svim drugim šalama iznosi 1 (tj. jedan redak item-to-item matrice sličnosti mora imati zbroj 1). Sličnosti šala dohvatite iz strukture `ItemSimilarity`.

Treći korak čini izrada hibridne matrice sličnosti koja kombinira vrijednosti normiranih sličnosti uz pripadajući težinski faktor. Rezultat trećeg koraka je jedna matrica s definiranom hibridnom sličnosti za parove šala. Hibridne matrice sličnosti se dobivaju korištenjem težinskih faktora tako da korisnik može definirati težinske udjele pojedine sličnosti prilikom izračuna hibridne matrice (u našem slučaju su dva težinska faktora za sličnost po sadržaju i sličnost po ocjenama korisnika). Hibridna sličnost se dobiva zbrajanjem dviju težinskih matrica sličnosti.

Nakon što ste napravili hibridnu matricu sličnosti, spremite je u odgovarajućem formatu da ju preporučitelj iz programskog okvira Apache Mahout može koristiti.

## Zadatak 3: Izrada hibridnog preporučitelja

(NAPOMENA: zadatak se temelji na hibridnoj matrici sličnosti iz prethodnog zadatka)

### Zadatak

Cilj zadatka je kreirati hibridni preporučitelj koristeći programski okvir Apache Mahout.

Uspješnim rješenjem ovog zadatka steći ćete sljedeća znanja:

- rad s programskim okvirom Apache Mahout

### Detaljni opis zadatka

Ovaj zadatak se temelji na rezultatu prethodnog zadatka, tj. potrebno je napraviti hibridni preporučitelj. Za podatkovni model iskoristite datoteku `jester_ratings.dat`, a za matricu sličnosti objekata koristite hibridnu matricu kreiranu u prošlom zadatku.

Također napravite evaluaciju kvalitete ova tri preporučitelja: hibridni i dva iz zadaće.

**IZVJEŠTAJ: Na sustav Moodle potrebno je predati izvještaj koji sadrži rješenja svih zadataka (odgovore na pitanja i pripadajući izvorni programski kod) do ponedjeljka 04.06.2018. u 09:00 sati**