

3. Domaća zadaća – ROVKP

Gregory Tomadin
JMBAG: 0036496597

1.Zadatak

1. U izlaznoj datoteci ima 11175 zapisa.
2. Šala najslićnija šali s ID-jem 100 je šala s ID-jem 138.
3. U obje šale se spominje predsjednik.
4. Ljudima se generalno sviđaju šale koje imaju sličan sadržaj, tako da bi preporuka po sadržaju trebala imati smisla.

2.Zadatak

1. Prvi preporučitelj: 103, 134, 139, 111, 63, 51, 141, 148, 137, 126
Drugi preporučitelj: 148, 129, 138, 132, 119, 144, 143, 127, 59, 58
2. Prvi preporučitelj ima bolju kvalitetu.
3. Za ove ulazne podatke je bolje koristiti mjeru log-likelihood.

3.Zadatak

1. Izvršeno je 9 MapReduce poslova.
2. Postoje razlike u izlaznim datotekama zbog implementacije algoritma koji izvršava recommenduserbased naredbe i zbog toga što koristimo threshold naredbu.