

Kombiniran vsebinsko-skupinski sistem za iskanje uporabniku prilagojenih multimedijskih vsebin

Matevž Kunaver, mentor: Jurij Tasič, Andrej Košir

*Fakulteta za elektrotehniko, Univerza v Ljubljani
E-pošta: matevz.kunaver@fe.uni-lj.si*

Povzetek. Uporabniku prilagojeno iskanje vsebin je postopek, pri katerem iz množice dostopnih vsebin izberemo tiste, za katere menimo, da jih bo uporabnik najbolj pozitivno sprejel (ocenil z visoko oceno). Zato potrebujemo podatke o uporabniku, ki jih dobimo s pomočjo analize njegovih preteklih dejanj – do katerih vsebin je dostopal in do katerih ne. Podatke, pridobljene s to analizo, nato shranimo v podatkovno strukturo, ki jo imenujemo uporabniški profil.

V okviru doktorskega dela smo razvili dva sistema za uporabniku prilagojene storitve ter jih testirali na dveh zelo različnih podatkovnih množicah. Prva podatkovna množica je bila dokaj prazna (vsebovala je samo majhen odstotek vseh mogočih ocen) ter je tako omogočila simuliranje uspešnosti sistema v okolju, kjer se je sistem šele začel uporabljati (torej se uporabniki prvič prijavljajo in podajajo ocene samo za ozek nabor vsebin). Druga podatkovna množica pa je bila skoraj polna (vnesene so bile skoraj vse ocene) in je tako omogočila simuliranje uspešnosti sistema v okolju, kjer se sistem uporablja že dalj časa in so uporabniki ocenili že večino vsebin ali pa okolju, kjer se zelo poredko dodajajo nove vsebine.

Sistema smo evalvirali s pomočjo izračuna vrednosti mere F , ki se na področju uporabniku prilagojenega iskanja vsebin najpogosteje pojavlja kot mera uspešnosti sistema. Poleg mere F smo naše vzorce tudi statistično analizirali, da smo lahko ugotovili, ali med posameznimi pristopi res obstajajo signifikantne razlike. Pri prvem sistemu smo se osredinili na področje skupinskega filtriranja in testirali dve različni metodi iskanja uporabnikov, da bi ugotovili, katera se bolje obnese (torej katera metoda uporabniku bolj zanesljivo predlaga vsebine). Ugotovili smo, da se najbolje obnese kombinacija obeh metod, ki tako deluje na širšem naboru uporabnikov. Poleg tega smo testirali tudi metode za izračun napovedane ocene, kjer smo preizkusili metodo utežene vsote in metodo Bayesovega približka. Na podlagi rezultatov smo razvili novo metodo, zasnovano na uteženi vsoti, ter jo uporabili v prvem sistemu.

Pri drugem sistemu smo obstoječe uporabniške modele razširili, jim dodali žanrske preference ter tako razvili hibridni sistem. Tako dobljeni sistem je deloval opazno hitreje (po nekaj sekund na uporabnika) in z večjo zanesljivostjo kot prvotno razviti sistem. S tem smo tudi pokazali, da kombinirani sistemi delujejo bolje kot sistemi, zasnovani samo na skupinskem ali samo na vsebinskem principu filtriranja vsebin, kar je razvidno iz primerjave potekov mer F posameznih sistemov.