

Društveni kontekst u preporučiteljskim sustavima

Darko Štriga

Sveučilište u Zagrebu Fakultet elektrotehnike i računarstva

Zavod za telekomunikacije

darko.striga@fer.hr

Sažetak — U radu su opisane dvije glavne skupine preporučiteljskih sustava: preporučitelji temeljeni na sadržaju i preporučitelji temeljeni na suradnji. Prikazani su osnovni mehanizmi preporučiteljskih sustava te su navedene njihove prednosti i nedostaci. U vidu poboljšanja preporučiteljskih sustava uveden je koncept društvenog konteksta. Društvene mreže danas kao jedna od najpopularnijih usluga na internetskoj infrastrukturi koja generira veliku količinu podataka predstavlja bogat izvor društvenog konteksta. Društveni kontekst predstavljaju međusobni odnos i interakcije korisnika na društvenim mrežama te prikazuju preferencije korisnika za određenu vrstu sadržaja. Na temelju razmatranja u radu zaključujemo da integracija društvenog konteksta u preporučiteljske sustave ima pozitivne efekte na njihovo funkcioniranje. Društvena preporuka sadržaja predstavlja personaliziranu informaciju jer sadržaj koji se preporučuje takvim inteligentnim sustavom preporučuje sadržaje na temelju ponašanja samoga korisnika i njegovog položaja u društvu tj. njegove interakcije s okruženjem.

Ključne riječi — društveni kontekst, društvene mreže, preporučiteljski sustavi, korisnički profili

I. UVOD

Suvremeni način života rezultira ogromnom količinom dostupnih podataka tako da se je pokazala potreba za što vremenski efikasnije, preciznije i personalizirano prikupljanje potrebnih podataka. Pritom, količina generiranih podataka bilježi eksponencijalni rast. Tako velika količina podataka najčešće korisnicima predstavlja veliki problem prilikom pronalaska željenog sadržaja. Rješenje za navedene probleme možemo pronaći u sustavima za preporučivanje koji analiziraju cjelokupni dostupni sadržaj i do korisnika filtriraju većinu sadržaja tako da korisnik u konačnici dobije samo sadržaj koji bi za njega mogao biti zanimljiv. Stoga možemo reći da su preporučiteljski sustavi programska podrška koja pruža vezu između velike količine dostupnog sadržaja i korisnikovih osobnih preferencija. Iako se korijeni preporučiteljskih sustava mogu pronaći u kognitivnoj znanosti, pretraživanju informacija, teorijama predviđanja i znanosti o menadžmentu, sustavi za preporučivanje su se u posljednjem desetljeću izdvojili kao neovisno područje istraživanja.

Ovisno o izvedbi, preporučiteljski sustavi brinu da do pojedinog korisnika dođu samo ponude koje su usko vezane

uz njegove interese. To se izvodi kontinuiranim praćenjem aktivnosti korisnika na:

- web-stranicama;
- društvenim mrežama (Facebook, Twitter, ...);
- platformama za internetsku televiziju (engl. *Internet protocol television, IPTV*);
- platformama za programe vjernosti (engl. *loyalty program*) i sl.

Na slici (Slika 1) je prikazana arhitektura sustava za preporuku. Komponenta *izvori informacija* predstavlja proces prikupljanja informacija o sadržaju za preporuku i korisnicima. Središnji dio preporučitelja je *sustav za preporuku* koji obuhvaća proces obrade prikupljenog sadržaja za preporuku i obradu profila korisnika. Posljednja komponenta preporučitelja je *korisnička aplikacija* koja omogućuje krajnjim korisnicima pregled preporučenog sadržaja.



Slika 1. Arhitektura sustava za preporuku

Profil korisnika je virtualna reprezentacija korisnika u informacijskom sustavu koja entitetima u sustavu omogućuje stvaranje modela korisnika. Ovisno o namjeni informacijskog sustava, korisnički profili se razlikuju po vrsti te opsegu informacija koje sadrže, ali i po tehnologiji koja je korištena za njihov zapis [1]. Profili korisnika kod preporučiteljskih sustava sadrže najčešće osobne podatke o korisniku i podatke o aktivnostima korisnika na ciljanoj platformi za preporuku. Sadržaj za preporuku su proizvodi i usluge koji zajedno čine skup za preporuku. *Profili sadržaja za preporuku* su profili koji

su generirani na temelju ključnih riječi tj. određenih specifičnosti promatranog sadržaja. Preporučeni sadržaj je podskup skupa sadržaja za preporuku koji je generiran na temelju korisničkog profila i profila sadržaja za preporuku.

U današnje vrijeme *društvene mreže* kao platforma više ne predstavljaju mjesto gdje je prosječan korisnik samo u interakciji sa svojim poznicima, već predstavljaju mjesto prikupljanja ciljanih informacija različitim pretplatama [2]. Pojam prijatelj na društvenim mrežama predstavlja povezanost između korisnika društvenih mreža. Ovisno o vrsti namjene društvene mreža povezanost može biti jednosmjerna i dvosmjerna. Vijesti i poruke objavljene putem društvenih stranica (Facebook stranice, *tweetovi* na društvenoj mreži Twitter,...) većina korisnika ne shvaća kao oglas za određenu marku (engl. *brand*), nego kao zanimljivu obavijest koju su voljni svojevrijedno pregledati, komentirati, označiti da im se taj sadržaj sviđa i podijeliti ga prijateljima [2]. Ovakvim ponašanjem krajnjih korisnika, gdje oni samostalno kreiraju profile i nesvjesno generiraju veliku količinu privatnih i osobnih podataka, omogućena je dostupnost korisničkim podacima putem društvenih mreža. Inteligentni sustavi analizom takvih podataka uspoređuju i rangiraju korisnike međusobno čime je moguće preciznije odrediti preferencije korisnika i odrediti zajednički interes za pojedine grupe korisnika što je ujedno i osnovni temelj za društvenu preporuku sadržaja. Tehnologije društvenih mreža omogućuju novu razinu sofisticiranosti kojom korisnik može brzo i direktno dobiti preporuke na temelju reakcije svojih prijatelja na određeni sadržaj u vidu komentara, recenzije i/ili ocjene. Ovakvim postupkom se ostvaruje viša razina osobnosti u preporuci u odnosu na preporuke koristeći uobičajene izvore kao što su novine ili web-portali.

Prednosti društvenih mreža uočile su mnoge kompanije. Ukoliko se promatra odnos uloženog i dobivenog, takva ponašanja korisnika na društvenim mrežama predstavljaju vrlo pogodno mjesto za vrlo efikasan način izgradnje imidža marke. Kako bi ubuduće kompanije mogle uspješno poslovati i plasirati svoje proizvode, nužno je stvaranje pozitivnog imidža na društvenim mrežama kao i okretanje iz pretežito tradicionalnih (televizijskih oglas, radio oglas, ...) prema suvremenom načinu promoviranja (internetsko oglašavanje, oglašavanje na pokretnim uređajima, ...) koji su se pokazali mnogo efikasniji [3][4][5][6]. Ipak, postoji problem. Iako su korisnici voljni pratiti tako generirane obavijesti, moguće je da pogrešnim pristupom i strategijom razviju negativan stav prema marki [2].

U poglavlju (II.) opisane su osnovne metode za izračunavanje preporuka i opisani su profili u preporučiteljskim sustavima. Poglavlje (III.) opisuje pojam i posljednje trendove na društvenim mrežama. Nadalje, poglavlje (IV.) daje pregled

primjene društvenog konteksta u preporučiteljskim sustavima te su u posljednjem poglavlju (V.) navedeni zaključci.

II. PREPORUČITELJSKI SUSTAVI

Preporučiteljski sustavi imaju zadatak korisnicima preporučivati sadržaj za koji su pokazali interes iz nekog skupa određenog sadržaja. Sadržaj koji se preporučuje su proizvodi (filmovi, glazba, ...) i usluge [7].

Jedan od najbolje znanih sustava za preporučivanje je preporučitelj internetske trgovine Amazona¹. Međutim postoji još mnogo drugih preporučiteljskih sustava na Internetu kao što su preporučitelji filmova MovieLens², Jinni³, IMDb⁴, Netflix⁵, preporučitelji glazbenog sadržaja LastFM⁶, Flokoon⁷, Musicoverly⁸, preporučitelji širokog spektra sadržaja kao što je TasteKid⁹ te razni drugi preporučitelji sadržaja.

Općenito, sustave za preporuku možemo podijeliti u dvije glavne kategorije [7]:

- preporučitelji temeljeni na sadržaju (engl. *content based*) i;
- preporučitelji temeljeni na suradnji (engl. *collaborative based*).

Kombiniranjem ovih dvaju pristupa dobivamo hibridne sustave za preporuku. Preporučitelji temeljeni na sadržaju imaju fokus na ocjenama korisnika i sličnosti između ocjenjenog sadržaja i ostalog dostupnog sadržaja za preporuku za izračunavanje preporuke za korisnika, dok klasični preporučiteljski sustavi temeljeni na suradnji koriste implicitne i eksplicitne ocjene prikupljene od drugih korisnika. Smislene preporuke za promatranog korisnika osiguraju se usporedbom povijesnih ocjena tj. ocjenjenog sadržaja drugih korisnika i ocjenjenog sadržaja promatranog korisnika [7].

Preporučitelj može na primjer [7]:

- predviđati ocjenu koju će korisnik dodijeliti preporučenom sadržaju (problem predviđanja – engl. *prediction problem*) ili;
- može iskoristiti predikciju za generiranje liste najboljih N sadržaja za preporuku (engl. *top-N recommendation problem*).

¹ <http://www.amazon.com/>

² <http://www.movielens.org>

³ <http://www.jinni.com/>

⁴ <http://www.imdb.com/>

⁵ www.netflix.com

⁶ <http://www.last.fm/>

⁷ <http://www.flokoon.com>

⁸ <http://musicoverly.com/>

⁹ <http://www.tastekid.com/>

Jedna od verzija preporučitelja zasnovanog na suradnji je *društveni preporučiteljski sustav* (engl. *social recommender system*) [7]. Za izgradnju sustava za društvenu preporuku koriste se aspekti i komponente tradicionalnog preporučitelja koji se dizajniraju korištenjem društvenih entiteta (subjekata). Koncept može uključivati slučaj gdje preporučeni sadržaj sam po sebi može biti društveni entitet osoba ili grupe osoba (npr. specijalne preporuke, engl. *expert-recommenders*) (primjer: preporuka mogućih prijatelja s kojima još nismo povezani na društvenoj mreži). Ipak, kod interpretacije fokus je većinom na korisnikove prijatelje koji ocjenjuju ili imaju stav o sadržaju za preporučivanje. Korisniku se preporučuje sadržaj od onih prijatelja koji po društvenom kontekstu imaju sličnost s korisnikovim ukusom i ponašanjem. Društveni kontekst može bit manje dinamičan (npr. prijatelji na društvenoj mreži Facebook) ili kratkotrajan i dinamičan s obzirom na situaciju u kojoj se korisnik nađe (npr. *ad-hoc* mreže raznih namjena) [8].

A. Profili u sustavima za preporuke

Profil je prezentiran skupom značajki koji se zovu atributi ili svojstva. Ovisno o namjeni imamo:

- profil sadržaja za preporuku i;
- profil korisnika sustava za preporuku.

Kada su svi sadržaji iz skupa sadržaja za preporuku opisani s istim skupom atributa i kada je taj skup atributa poznat tada možemo reći da je sadržaj za preporuku predstavljen pomoću strukturiranih podataka tj. strukturiranih profila. U takvim slučajevima mogu se koristiti mnogi algoritmi za strojno učenje za učenje tj. kreiranje i analizu profila sadržaja za preporuku [9]. U većini prijašnjih sustava za preporuku korisnički profili su strukturirani, ali pojavom društvenih mreža i dostupnošću velike količine generiranih podataka te sve većom popularnošću ne-SQL (engl. *structured query language*) baza podataka, koje omogućuje lagano baratanje s nestrukturiranim podacima, korisnički profili gube svoju strukturu.

U mnogim preporučiteljima temeljenim na sadržaju opisi sadržaja su tekstualnog formata najčešće dohvaćenih s internetskih stranica, elektroničke pošte, internetskih članaka ili iz samog opisa sadržaja za preporuku. Za razliku od strukturiranih podataka atributi koji opisuju takvu vrstu sadržaj nisu dobro definirane vrijednosti tako da imamo nestrukturirane profile sadržaja za preporuku. Ovakva vrsta nestrukturiranog opisa sadržaja kreira probleme kod učenja profila sadržaja za preporuku zbog dvosmislenosti prirodnog jezika. Problem je tradicionalan kod profila temeljenih na ključnim riječima kod kojih nije moguće raspoznati semantiku korisničkog interesa jer algoritmi najčešće uspoređuju podudaranje samih nizova znakova. Primjerice ako je niz znakova ili neka njegova morfološka varijanta pronađena u

profilu korisnika i u sadržaju za preporuku tada je zabilježeno takvo podudaranje i takav sadržaj se smatra kao relevantan.

Uspoređivanje nizova znakova teksta pati od problema:

- višeznačnost – jedna riječ ima više značenja te;
- sinonimi – više riječi imaju isto značenje.

Posljedica prethodno navedenih problema je ta da je moguće zbog sinonima propustiti relevantne informacije za korisnika, ako profil korisnika ne sadrži točne riječi iz promatranog opisa sadržaja, dok se zbog višeznačnosti pogrešan sadržaj može smatrati relevantnim. Semantička analiza i njezina integracija u model personalizacije je jedan od inovativnijih i najzanimljivijih pristupa rješavanju navedenih problema. Ključna ideja takvih algoritama je usvajanje baza znanja, kao što su leksikoni ili ontologije, za označavanje sadržaja za preporuku i predstavljanje profila kako bi se dobila semantička interpretacija o sadržaju koje korisnik želi [10].

B. Preporučitelji temeljeni na sadržaju

Preporučitelji temeljeni na sadržaju (engl. *content-based filtering*) oslanjaju se na pretpostavku da će se korisniku svidjeti sadržaji koji imaju slične karakteristike sadržajima koji su mu se svidjeli u prošlosti. Sustav na temelju prethodno ocjenjenog i/ili odabranog sadržaja analizira i generira preporuke koje se temelje na sličnim karakterističnim elementima u sadržajima kod pozitivno rangiranog sadržaja. Takvim se postupkom izgrađuje profil korisnika i nastoji se shvatiti što korisnik preferira. Karakteristični elementi za određeni ocjenjeni sadržaj se uspoređuju s karakterističnim elementima kategorija i skupova sadržaja te se izdvajaju oni sadržaji koji se po karakterističnim elementima u najvećoj mjeri podudaraju s preferencijama korisnika. Na taj se način značajno smanjuje broj potencijalnog zanimljivog sadržaja za preporuku određenom korisniku.

Preporučitelji temeljeni na sadržaju imaju nekoliko prednosti u odnosu na preporučitelje temeljene na suradnji [10]:

- **Neovisnost korisnika** – Preporučitelji temeljeni na sadržaju iskorištavaju isključivo ocjene pojedinog korisnika za izgradnju njegovog profila, dok sustavi temeljeni na suradnji imaju potrebu za ocjenama od drugih korisnika kako bi pronašli njegove najbližije korisnike (engl. *nearest neighbors*). Takvi slični korisnici imaju sličan ukus koji je određen na temelju približno jednako dodijeljenih ocjena za ocjenjeni sadržaj.
- **Transparentnost** – Objašnjenje kako je preporučitelj temeljen na sadržaju generirano listu preporuka za korisnika moguće je dobiti eksplicitno na temelju liste značajki koje su uzrokovale da se neki sadržaj nađe u preporukama. Ova značajka omogućuje korisniku da se odluči da li vjerovati takvoj preporuci

ili ne. Sustavi temeljeni na suradnji su crne kutije kod kojih je jedino objašnjenje zašto se je neki sadržaj našao u preporukama je taj da neki nepoznati korisnik ima sličan ukus kao i promatrani korisnik te da je on označio da mu se taj sadržaj sviđa.

- **Novi sadržaj** – Preporučitelji temeljeni na sadržaju su sposobni preporučivati sadržaje koji nisu još ocijenjeni ni od bilo kojeg korisnika. Posljedica toga je da se izbjegava hladan start za novi sadržaj koji se u ovome slučaju javlja kod preporučitelja temeljene na suradnji.

Ipak, preporučitelji temeljeni na sadržaju imaju nekoliko nedostataka [10]:

- **Ograničeni opis sadržaja** – Tehnike preporučitelja temeljenog na sadržaju imaju prirodno ograničenje u vrsti sadržaja koji mogu preporučivati. Često uspješnost rada sustava za preporuke temeljenih na sadržaju ovise o dostupnim podacima i domenskom znanju kojima se opisuje sadržaj (npr. za preporučivanje filmova preporučitelj mora znati glumce, direktore, opis radnje i sl.). Potrebno je iz tog razloga omogućiti automatsko opisivanje sadržaja ili dodijeliti ručno karakteristične elemente sadržajima. Kod većine domena automatsko opisivanje sadržaja je vrlo složeno kao npr. kod opisa višemedijskih sadržaja, dok je ručni opis vremenski zahtjevan.
- **Pretjerana specijalizacija** – Preporučitelji temeljeni na sadržaju ne mogu preporučiti neočekivani sadržaj. Zbog načina određivanja vrijednosti funkcije korisnosti preporučitelj favorizira sadržaj čija je ocjena vrlo visoka (rangirana) te preporučuje sadržaj kada se karakteristični elementi sadržaja podudare s profilom korisnika. Posljedica toga je da se korisniku nikada ne preporučuju sadržaji koji se po karakterističnim elementima razlikuju od onoga što je korisnik ranije označio da mu se sviđa, već se nude međusobno vrlo slični sadržaji. Ova mana označuje tendenciju da preporučiteljski sustav producira preporuke s ograničenim stupnjem novosti (engl. *serendipity problem*). Primjer ove mane u praksi je kada korisnik ocjenjuje filmove samo kojima je bio direktor/producent Stanley Kubrick¹⁰ tada će korisnik dobivati samo takav tip filmova za preporuku. „Savršeni“ preporučiteljski sustav temeljen na sadržaju će vrlo rijetko pronaći nešto novo čime se ograničava raspon primjene za što se može koristiti.
- **Novi korisnik** – Naime, zbog načina rada preporučitelja temeljenog na sadržaju nužno je da korisnik ocjeni dovoljan broj sadržaja kako bi sustav mogao stvarno razumjeti preferencije korisnika i

generirati mu odgovarajuće preporuke. Dakle, kada su za novog korisnika dostupne samo nekoliko ocjena preporučitelj nije u stanju osigurati vjerodostojne preporuke.

1) Metoda vektorski prostorni model temeljen na ključnim riječima

Najveći broj preporučitelja temeljenih na sadržaju najčešće koriste [10]:

- relativno jednostavne modele za izračunavanje podudaranja ključnih riječi (primjer: usporedba stringova) ili;
- vektorski prostorni model (engl. *vector space model*, VSM) s osnovnim *TF – IDF* težinama (engl. *term frequency–inverse document frequency*)

VSM temeljen na ključnim riječima (engl. *keyword-based vector space model*) je prostorni prikaz tekstualnog dokumenta. Ključne riječi ili pojmovi u ovome modelu predstavljaju karakteristične oznake tj. riječi koje opisuju vjerodostojno profil sadržaja za preporuku. U modelu je svaki tekstualni dokument (u našem slučaju profil sadržaja za preporuku) predstavljen s vektorom u n -dimenzionalnom prostoru, gdje svaka dimenzija predstavlja pojam iz sveukupnog rječnika kolekcije dokumenata. Formalno, svaki dokument je predstavljen kao težinski vektor pojmova, gdje svaka težina ukazuje na stupanj povezanosti između dokumenta i pojma. Neka $D = \{d_1, d_2, \dots, d_N\}$ predstavlja skup dokumenata i $T = \{t_1, t_2, \dots, t_n\}$ predstavlja skup pojmova u rječniku. Svaki dokument d_j je predstavljen kao vektor u n -dimenzionalnom vektorskom prostoru kao $d_j = \{w_{1j}, w_{2j}, \dots, w_{nj}\}$, gdje je w_{kj} težina za pojam t_k u dokumentu d_j [10].

Najpoznatiji model određivanja težinskog faktora w_{ij} je *TF – IDF* model. Pretpostavke modela [10][11]:

- rijetki pojmovi nisu manje relevantni od čestih pojmova (*IDF* pretpostavka);
- višestruka pojava pojma u dokumentu nije manje relevantna od jedne pojave (*TF* pretpostavka);
- dugi dokumenti nisu preferirani nasuprot kratkih dokumenata (pretpostavka normalizacije).

Drugim riječima, pojmovi koji se često pojavljuju u jednom dokumentu (*TF = term-frequency*), ali rijetko u drugim dokumentima u skupu (*IDF = inverse-document-frequency*) su najvjerojatnije relevantni za temu dokumenta. Normalizacijom težinskog vektora eliminira se prednost dužih dokumenata koji imaju veću vjerojatnost da sadrže određeni pojam.

¹⁰ <http://www.imdb.com/name/nm0000040/>

Ove pretpostavke su definirane funkcijom $TF - IDF$ [10]:

$$TF - IDF(t_k, d_j) = \underbrace{TF(t_k, d_j)}_{TF} * \underbrace{\log \frac{N}{n_k}}_{IDF} \quad (1)$$

$$TF(t_k, d_j) = \frac{f_{k,j}}{\max_z f_{z,j}} \quad (2)$$

U tablici (Tablica 1) su objašnjene oznake korištene u ovome poglavlju.

Tablica 1. Objašnjenje oznaka

Oznaka	Objašnjenje
N	broj dokumenata u kolekciji
n_k	broj dokumenata u kolekciji u kojima se pojam t_k pojavljuje najmanje jednom
t_k	pojam u rječniku
d_k	dokument u skupu
$f_{k,j}$	frekvencije pojavljivanja $f_{k,j}$ svih pojmova t_k u dokumentu d_j
$f_{z,j}$	frekvencije pojavljivanja $f_{z,j}$ svih pojmova t_z u dokumentu d_j
$\max_z f_{z,j}$	Funkcija maksimum ($\max_z f_{z,j}$) u dokumentu izračunat je preko frekvencije pojavljivanja $f_{z,j}$ svih pojmova t_z u dokumentu d_j .
T	Broj pojmova u dokumentu

Kako bi dobiveni rezultat indeks težine (3) preslikali u interval $[0, 1]$ potrebno je provesti normalizaciju. Dobiveni indeks težine se najčešće normalizira pomoću kosinus normalizacije da bi se ispunila pretpostavka normalizacije [10].

$$w_{k,j} = \frac{TF - IDF(t_k, d_j)}{\sqrt{\sum_{s=1}^{|T|} (TF - IDF(t_s, d_j))^2}} \quad (3)$$

Kao što je ranije navedeno, mjera sličnosti je potrebna kako bi se utvrdila „bliskost“ između dvaju dokumenata te je za to najraširenija u primjeni kosinus sličnost (4) [10].

$$sim(d_i, d_j) = \frac{\sum_k w_{ki} * w_{kj}}{\sqrt{\sum_k w_{ki}^2} * \sqrt{\sum_k w_{kj}^2}} \quad (4)$$

Kod preporučitelja temeljenog na VSM, korisnički profil i sadržaj za preporuku su zastupljeni težinskim vektorima. Predviđanje korisnikovog interesa za određeni sadržaj može se izvesti izračunom kosinus sličnosti tih dvaju vektora.

2) Opisani postupak primjene metode vektorskog prostornog model temeljenog na ključnim riječima

U nastavku je prikazan način izračuna sličnosti između profila sadržaja za preporuku korištenjem metode vektorskog prostornog model temeljenog na ključnim riječima. Profili sadržaja za preporuku su profili filmova s internetske baze filmova IMDb.

U prvome stupcu tablice (Tablica 2) nalazi se skup pojmova u rječniku T . Rječnik sadrži pojmove koji opisuju vrstu žanrova i glumce. Prvi redak u tablici predstavlja skup dokumenata D . Dokumenti predstavljaju profile sadržaja za preporuku tj. filmove. Izračunate vrijednosti u tablici predstavljaju vrijednost TF koja je izračunata ne temelju formule (2).

Tablica 2. VSM primjer: prikaz izračunatih vrijednosti TF

	Safe House ¹¹	Flight ¹²	The Bourne Ultimatum ¹³
Action	1	0	1
Crime	1	0	1
Mystery	1	0	0
Drama	0	1	0
Thriller	0	0	1
Matt Damon	0	0	1
Denzel Washington	1	1	0

Vrijednosti u tablici (Tablica 3) predstavljaju vrijednosti $TF - IDF(t_k, d_j)$ izračunate na temelju formule (1).

Tablica 3. VSM primjer: prikaz izračunatih vrijednosti TF-IDF

$TF - IDF(t_k, d_j)$	Safe House	Flight	The Bourne Ultimatum
Action	1*0,176	0	1*0,176
Crime	1*0,176	0	1*0,176
Mystery	1*0,477	0	0
Drama	0	1*0,477	0
Thriller	0	0	1*0,477
Matt Damon	0	0	1*0,477
Denzel Washington	1*0,176	1*0,176	0

¹¹ <http://www.imdb.com/title/tt1599348/>

¹² <http://www.imdb.com/title/tt1907668/>

¹³ <http://www.imdb.com/title/tt0440963/>

Tablica 4. VSM primjer: međurezultati

	Safe House	Flight	The Bourne Ultimatum
$\sqrt{\sum_{s=1}^{ T } TF - IDF(t_s, d_j)^2}$	0,566	0,508	0,719
$\sqrt{\sum_k w_{ki}^2}$	0,566	0,508	0,719

U tablici (Tablica 5) su izračunate vrijednosti težina za svaki pojam u svakom profilu sadržaja za preporuku koje se izračunavaju formulom (3) uz prikaz pojedinih međurezultata u tablici (Tablica 4).

Tablica 5. VSM primjer: težine pojedinih pojmova u profilu sadržaja za preporuku

$w_{k,j}$	Safe House	Flight	The Bourne Ultimatum
Action	0,311	0	0,245
Crime	0,311	0	0,245
Mystery	0,843	0	0
Drama	0	0,939	0
Thriller	0	0	0,663
Matt Damon	0	0	0,663
Denzel Washington	0,311	0,346	0

Posljednji korak kod analizirane metode predstavlja izračunavanje kosinus sličnosti na temelju izračunatih vrijednosti težina (Tablica 5) korištenjem formule (4).

Tablica 6. VSM primjer: izračunate sličnosti između filmova

	Safe House	Flight	The Bourne Ultimatum
Safe House	X	0,374	0,374
Flight	0,374	X	0
The Bourne Ultimatum	0,374	0	X

Na temelju izračunate sličnosti između filmova zaključujemo da je profil film *Safe House* po odabranim pojmovima podjednako sličan profilima filmova *Flight* i *The Bourne*, dok profili filmova *Flight* i *The Bourne* nemaju zajedničkih pojmova.

C. Preporučitelji temeljeni na suradnji

Preporučitelji temeljeni na suradnji (engl. *collaborative recommender*) temelje se na sakupljanju i analizi velike količine podataka o ponašanju, aktivnostima i preferiranom sadržaju korisnika. Glavna ideja preporučitelja temeljenog na suradnji je iskoristiti informacije o prošlim ponašanjima ili izraženim mišljenjima korisnika za predviđanje koji sadržaj će se korisniku najvjerojatnije svidjeti i za koji će biti

zainteresiran [12]. Danas su ovakve vrste preporučitelja u širokoj industrijskoj upotrebi naročito kao preporučitelji u internetskim trgovinama, gdje se pomoću njih provodi promocija sadržaja i time povećava prodaja. Metode preporučitelja temeljenog na suradnji preporučuju korisniku specifičan sadržaj temeljen na uzorcima ocjena ili upotrebe (npr. kupnja) bez potrebe o egzaktnim informacijama o korisniku ili sadržaju. Takvi preporučiteljski sustavi oslanjaju se na razne vrste ulaza, a jedan od najvažnijih ulaza je direktno izražavanje interesa korisnika za određeni sadržaj kroz razne oblike ocjena. Takve ocjene koje su korisnici dodijelili sadržajima koriste se kao aproksimativna reprezentacija njihovih interesa i potreba. Primjerice Netflix¹⁴ sakuplja ocjene zvjezdice za filmove, dok korisnici TiVo¹⁵ prikazuju svoje sklonosti za televizijske emisije pritiskom odgovarajućih tipaka na televizijskom upravljaču. Iz razloga što eksplicitna povratna veza nije uvijek dostupna, neki preporučiteljski sustavi izvode zaključke o preferencijama korisnika iz više dostupnijih implicitnih povratnih veza. Implicitne povratne veze neizravno izražavaju mišljenje korisnika promatranjem ponašanja korisnika (npr. korisnik koji kupuje mnogo knjiga od istoga autora najvjerojatnije voli knjige tog autora). Vrste implicitnih povratnih veza uključuju [10]:

- povijest kupnje;
- povijest pretraživanja weba;
- povijest pregledavanja weba te čak i;
- povijest kretanja miša.

Osnovni preporučitelj temeljen na suradnji uzima kao jedini ulaz matricu korisnik-sadržaj popunjenu ocjenama i uobičajeno kreira dvije vrste preporuka [12]:

- numerička vrijednost koja označava u kojem stupnju će se određeni sadržaj sviđati ili ne određenom korisniku;
- lista n preporuka sadržaja

1) Preporuka sadržaja na temelju sličnih korisnika

Glavna ideja preporuke sadržaja na temelju sličnih korisnika (engl. *User-based nearest neighbor recommendation*) je da se na temelju ocjenjenog sadržaja korisnika identificiraju ostali korisnici koji imaju slične preferencije. Tada se za svaki sadržaj koji korisnik nije još vidio izračunava vjerojatnost, koliko će mu se sadržaj sviđati, na temelju ocjena koje su dodijelili njemu slični korisnici po preferencijama. Temeljne pretpostavke takvog modela su da [12]:

- ako su korisnici imali slične ukuse u prošlosti, da će imati slične ukuse i u budućnosti te;
- korisničke preferencije su konzistentne tijekom vremena tj. ukus korisnika se s vremenom ne mijenja.

¹⁴ <https://signup.netflix.com/global>

¹⁵ <http://www.tivo.com/>

Kod analize ove metode koristit će se oznake navedene u tablici (Tablica 7).

Tablica 7. Objasnjenje oznaka koje će se koristiti u analizi

Oznaka	Opis
n	Broj korisnika
m	Broj sadržaja za preporuku
$U = \{u_1, \dots, u_n\}$	Skup korisnika
$P = \{p_1, \dots, p_m\}$	Skup sadržaja za preporuku
$R(n \times m)$	Matrica ocjena R dimenzija $n \times m$, gdje je $r_{i,j}$ ocjena, a $i \in 1 \dots n, j \in 1 \dots m$.

Za određivanje sličnosti korisnika vrlo česta mjera korištena u preporučiteljskim sustavima je Pearsonov koeficijent korelacije. Pearsonov koeficijent korelacije koristi se u slučajevima kada između varijabli postoji linearna povezanost. U skladu s veličinom ovoga koeficijenta može se odrediti smjer i jačina linearne povezanosti među promatranim varijablama. Vrijednost Pearsonovog koeficijenta korelacije kreće se od +1 (savršena pozitivna korelacija) do -1 (savršena negativna korelacija). Predznak koeficijenta upućuje na smjer povezanosti između varijabli, tj. je li veza pozitivna ili negativna [13]. Pearsonov koeficijenta korelacije tj. sličnost $sim(a, b)$, korisnika a i korisnika b , određena matricom ocjena R je definirana formulom (5). Simbol \bar{r}_a označava prosječnu ocjenu svih ocjena ocjenjenog sadržaja korisnika a [12].

$$sim(a, b) = \frac{\sum_{p \in P} (r_{a,p} - \bar{r}_a) * (r_{b,p} - \bar{r}_b)}{\sqrt{\sum_{p \in P} (r_{a,p} - \bar{r}_a)^2} * \sqrt{\sum_{p \in P} (r_{b,p} - \bar{r}_b)^2}} \quad (5)$$

Za sadržaj koji korisnik nije pregledao, a njegovi „slični“ prijatelji jesu predviđamo vrijednost ocjene koje bi korisnik a dao za sadržaj p (formula 6) [12]. Oznaka N označava skup najbližih (sličnih) prijatelja.

$$pred(a, p) = \bar{r}_a + \frac{\sum_{b \in N} sim(a, b) * (r_{b,p} - \bar{r}_b)}{\sum_{b \in N} sim(a, b)} \quad (6)$$

U prethodnom primjeru za izračun sličnosti između korisnika je korišten Pearsonov koeficijent korelacije, u literaturi se još spominju metrike kao što su [12]:

- korigirana kosinus sličnost (engl. *adjusted cosine similarity*);
- Spearmanov rang koeficijent korelacije (engl. *Spearman's rank correlation coefficient*) te;
- srednja kvadratna pogreška (engl. *mean squared error*).

Empirijske analize pokazuju da je za izračun sličnosti za model preporuke najbližih susjeda temeljene na korisnicima najbolje rezultate pokazao model s Pearsonov koeficijent korelacije

[14]. Međutim, kod preporučiteljskih modela temeljenih na sadržaju je uočeno da mjera kosinus sličnosti nadmašuje model s Pearsonovim koeficijentom korelacije [14].

Ipak, za pronalazak najbližih korisnika korištenjem samo tehnike Pearsonovog koeficijenta korelacije i ponderiranjem tih ocjena ne može biti najbolji izbor. Razmotrimo činjenicu, da u većini domena postoje neke stvari koje preferiraju svi korisnici. Stoga, mjera kao Pearsonov koeficijent korelacije neće uzeti u obzir da sadržaj koji je općenito označen kao popularan ima manji značaj nego sadržaj koji su specifično označili manji skup korisnika da im se sviđa. Kao posljedica toga je u radu [15] predloženo uvođene transformacijske funkcije na ocjenjeni sadržaj. Transformacijska funkcija ima zadatak da reducira relativnu važnost općenito popularnog sadržaja. Pokazano je da u slučaju kada predviđanje temeljimo na ocjenama sličnih korisnika koji imaju samo mali broj zajedničko ocjenjenog sadržaja je loš izbor i dovodi do lošeg predviđanja [14].

Za izračunavanje preporuka uzima se u obzir samo dio korisnika iz skupa korisnika. Ukoliko bi uključili sve korisnike, to ne bi samo negativno utjecalo na preporuke nego bi i značajno narušilo performanse (vrijeme izračuna). Uobičajene tehnike za smanjenje skupa sličnih korisnika se definira određivanjem minimalnog praga sličnosti ili ograničavanjem skup sličnih korisnika na fiksni broj k . Potencijalni problem ovih dvaju tehnika opisan je u radu [16]. Ukoliko je prag sličnosti između korisnika prevelik tada se dobije vrlo mali broj sličnih korisnika. Posljedica toga je da za velik dio sadržaj za preporuku nije moguće provesti predviđanje (smanjena pokrivenost). U suprotnosti, kada je prag sličnosti prenisko tada broj sličnih korisnika nije značajno reducirani.

Vrijednost odabrana za k -broj elemenata u skupu sličnih korisnika ne utječe na pokrivenost. Međutim, problem pronalaženja „dobre“ vrijednosti za k još uvijek postoji: kada je broj sličnih korisnika k prevelik tada preveliki broj korisnika s neznačajnom sličnosti donose šum u preciznost preporuke. Kada je broj k premalen (npr. manji od 10) tada utjecaj na kvalitetu predviđanja može imati negativan utjecaj [14]. Analiza skupa podataka MovieLens¹⁶ pokazala je da se razuman broj sličnih korisnika kreće u rasponu od 20 do 50 [17].

Slijedeći primjer opisuje primjenu algoritma promatrane metode. U tablici (Tablica 8) se nalaze dodijeljene ocjene korisnika za sadržaj (filmove). Vrijednost ocjena su u intervalu od 1 do 5, gdje ocjena 5 označuje da se korisniku jako sviđa sadržaj. Glavna točka razmatranja je korisnik *Matija* koji nije ocijenio film *Side Effects* te nam je zadatak odrediti da li će se film svidjeti *Matiji* ili ne. Ukoliko možemo predvidjeti hoće li

¹⁶ <http://www.movielens.org/>

se film svidjeti *Matiji* tada možemo uključiti taj film na njegovu listu preporuka. Za potrebe predviđanja cilj nam je pronaći korisnike koji imaju sličan ukus kao i *Matija* i tada uzeti ocjene te grupe korisnika za promatrani film (*Side Effects*) te odrediti da li će se film svidjeti i *Matiji*.

Tablica 8. Baza ocjenjenog sadržaja za primjere preporučitelja temeljenog na suradnji

	Safe House	Flight	The Bourne Ultimatum	Side Effects ¹⁷
Matija	5	4	5	?
Ana	5	3	4	3
Ines	2	3	5	4
Filip	4	3	4	3

U tablici (Tablica 9) se nalaze izračunate prosječne ocjene dodijeljene od strane korisnika.

Tablica 9. Primjer preporuke sadržaja na temelju sličnih korisnika: prosječna ocjena dodijeljena od strane korisnika

Matija	Ana	Ines	Filip
4.67	3.75	3.50	3.50

Korištenjem formule (5) u tablici (Tablica 10) izračunate su sličnosti između korisnika za kojega je potrebno odrediti da li će mu se film svidjeti i preostalih korisnika. Na temelju dobivenih rezultata korisnici *Ana* i *Filip* imaju slične ukuse kao i *Matija* te oni ulaze u skup sličnih korisnika.

Tablica 10. Primjer preporuke sadržaja na temelju sličnih korisnika: izračunate sličnosti korisnika

	Ana	Ines	Filip
Matija	0.83	0.19	0.94

Korištenjem formule (6) i promatranjem samo korisnika iz skupa sličnih korisnika (*Ana*, *Filip*) dobivamo (7):

$$\begin{aligned} \text{pred}(\text{Matija}, \text{Side Effects}) &= \\ &= 4.67 + \frac{0.83 * (3 - 3.75) + 0.94 * (3 - 3.5)}{0.83 + 0.94} \quad (7) \\ &= 4.05 \end{aligned}$$

Dobivena vrijednost predstavlja najvjerojatniju ocjenu koju bi korisnik *Matija* dodijelio za film *Side Effects*. Vrijednost 4.05 označuje da postoji velika vjerojatnost da će se film svidjeti korisniku te na temelju tog zaključka film uvrštavamo na listu preporuka.

2) Preporuka sadržaja na temelju slično ocjenjenog sadržaja

Primjena preporučitelja sadržaja temeljenog na sličnim korisnicima (engl. *Item-based nearest neighbor recommendation*) na komercijalnim internetskim stranicama kada postoji nekoliko milijuna korisnika i nekoliko milijuna sadržaja za preporuku susrećemo se s velikim izazovima. Konkretno potrebno je pretražiti veliki broj potencijalnih sličnih korisnika što čini izračun nemoguć u stvarnom vremenu. Stoga velike komercijalne internetske stranice često koriste preporučitelje temeljene na sadržaju koji su pogodniji za ne-stvarnovremensku pripremu preporuka. Takvim postupkom se omogućuje izračun preporuka za korisnike u stvarnom vremenu čak i za velike matrice ocjena [18][19].

Glavna ideja algoritma za izračun *preporuka sadržaja na temelju slično ocjenjenog sadržaja* (engl. *Item-based nearest neighbor recommendation*) je izračunati preporuke pomoću sličnosti između sadržaja za preporuku, a ne sličnosti između korisnika. Kod ovog pristupa kosinus sličnosti je uspostavljen kao standardna metrika. Metoda računanja kosinus sličnosti je slična kao i kod sustava preporučivanja temeljenog na sadržaju, gdje se sličnost mjeri između vektora $TF - IDF$ težina (dobivenih na temelju ključnih riječi koje opisuju profil sadržaja za preporuku), dok se kod sustava preporučivanja temeljenog na suradnji mjeri između vektora ocjena koje su korisnici dali sadržaju. Kako bi smanjili vrijeme potrebno za računanje preporuka, određeni sustavi za preporučivanje periodički izračunavaju sve sličnosti. Na taj način ovi sustavi brže poslužuju korisničke preporuke koristeći unaprijed proračunate sličnosti između sadržaja za preporuku.

Sličnost između dva sadržaja a i b , gledani kao vektori ocjena \vec{a} i \vec{b} , je definirana (8) [12]:

$$\text{sim}(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| * |\vec{b}|} \quad (8)$$

Simbol \cdot označava skalarni umnožak vektora, a $|\vec{a}|$ je euklidska udaljenost vektora. Moguće vrijednosti su u rasponu od 0 do 1. Osnovna kosinus mjera ne uzima u obzir prosječnu ocjenu ponašanja korisnika. Taj problem je riješen korištenjem korigirane kosinusne mjere koja uzima u obzir prosječnu ocjenu ponašanja korisnika. Nova korigirana vrijednost ocjene sadržaja korisnika se dobije na način da se dodijeljenoj ocjena za pojedini sadržaj oduzme prosječna ocjena svih dodijeljenih ocjena koje je korisnik dodijelio. Vrijednost korigirane kosinus mjere je u intervalu od -1 do 1.

¹⁷ <http://www.imdb.com/title/tt2053463/>

Neka je U skup korisnika koji su ocijenili oba sadržaja a i b te je tada korigirana kosinus mjera dana formulom (9) [12]:

$$sim(a, b) = \frac{\sum_{u \in U} (r_{u,a} - \bar{r}_u) * (r_{u,b} - \bar{r}_u)}{\sqrt{\sum_{u \in U} (r_{u,a} - \bar{r}_u)^2} * \sqrt{\sum_{u \in U} (r_{u,b} - \bar{r}_u)^2}} \quad (9)$$

Nakon izračuna sličnosti između sadržaja za preporuku, sada možemo predviđati ocjenu za sadržaj p koji korisnik u nije pogledao (10) [12]:

$$pred(u, p) = \frac{\sum_{i \in \text{ocijenjeniSadržaj}(u)} sim(i, p) * r_{u,i}}{\sum_{i \in \text{ocijenjeniSadržaj}(u)} sim(i, p)} \quad (10)$$

Kao što i kod prethodno analizirane metode broj razmatranih sličnih korisnika je ograničen na određeni broj tj. za izračun preporuka se uzima samo određen broj najsličnije ocjenjenog sadržaja.

Slijedeći primjer opisuje primjenu algoritma promatrane metode. Za ovaj primjer koristimo vrijednosti ocjena iz tablice (Tablica 8). Najprije izračunavamo kosinus sličnosti između dva filma korištenjem formule (8):

$$sim(\text{The Bourne Ultimatum}, \text{Side Effects}) = \frac{4 * 3 + 5 * 4 + 4 * 3}{\sqrt{4^2 + 5^2 + 4^2} * \sqrt{3^2 + 4^2 + 3^2}} = 0,9996 \quad (11)$$

Tablica 11. Primjer preporuke sadržaja na temelju slično ocjenjenog sadržaja: prosječna ocjena ocjenjenog sadržaja

Safe House	Flight	The Bourne Ultimatum	Side Effects
4.00	3.25	4.5	3.33

S obzirom da kosinus sličnosti ne uzima u obzir prosječnu ocjenu ponašanja korisnika tada izračunavamo korigiranu kosinus mjeru formulom (9). U ovom postupku transformiramo originalne ocjene iz tablice (Tablica 8) oduzimanjem izračunatih vrijednosti iz tablice (Tablica 11):

Tablica 12. Primjer preporuke sadržaja na temelju slično ocjenjenog sadržaja: korigirane vrijednosti ocjena

	Safe House	Flight	The Bourne Ultimatum	Side Effects ¹⁸
Matija	0.33	-0.67	0.33	?
Ana	1.25	-0.75	0.25	-0.75
Ines	-1.50	-0.50	1.50	0.50
Filip	0.50	-0.50	0.50	-0.50

¹⁸ <http://www.imdb.com/title/tt2053463/>

Izračunate vrijednosti iz tablice (Tablica 12) uvrstimo u formulu (8):

Tablica 13. Primjer preporuke sadržaja na temelju slično ocjenjenog sadržaja: izračunate sličnosti između filmova

	Safe House	Flight	The Bourne Ultimatum
Side Effects	-0.93	0.53	0.19

Na temelju dobivenih rezultata (Tablica 13) jedino film *Flight* stavljamo u skup sličnih filmova filmu *Side Effects* jer je njegova sličnost najveća (0.53). Nadalje slijedi određivanje predikcije uvrštavanjem podataka u formulu (10) te dobivamo (12):

$$pred(\text{Matija}, \text{Side Effects}) = \frac{0.53 * 4}{0.53} = 4 \quad (12)$$

Izračunata vrijednost predstavlja najvjerojatniju ocjenu koju bi korisnik *Matija* dodijelio za film *Side Effects*. Ocjena 4 označuje da postoji velika vjerojatnost da će se film svidjeti korisniku te na temelju tog zaključka film uvrštavamo na listu preporuka.

III. DRUŠTVENE MREŽE

Danas su društvene mreže općeprihvaćeno sredstvo interakcije s poznanicima i okruženjem gdje 60 posto odraslog svjetskog stanovništva koristi društvene mreže i čak 91 posto odraslog stanovništva koji se koristi Internetom. Društvene mreže su tako postale najpopularnija web-aktivnost gdje korisnici provedu 22 posto svojeg vremena na Internetu, dok je slijedeća aktivnost pretraživanje sadržaja s 21 posto, čitanje sadržaja 20 posto, komunikacija elektroničkom poštom i stvarnovremenskim porukama 19 posto [20][21][22][23]. Promatrajući odnose prosječnog mjesečnog utrošenog vremena čovjeka iz 2006. i 2011. godine uočljivi su promjene u trendovima. Tako su zabilježeni rast provedenog vremena na društvenom mrežama i gledanju televizije na Internetu, dok je kod svih ostalih aktivnosti zabilježen pad (Tablica 14) [22].

Tablica 14. Prosječno mjesečno vrijeme provedeno na aktivnostima [3]

Aktivnosti	2006.	2011.	Trend
Društvene mreže	2.7h	6.9h	+155.6%
Telefon, e-mail, pošta	5.7h	4.8h	-15.8%
Komunikacija uživo	22.8h	21h	-7.9%
Briga o ukućanima	15.9h	15.3h	-3.8%
Gledanje televizije putem klasičnih prijarnika	71.1h	59.4h	-16.5%
Gledanje televizije na Internetu	6.3h	23.1h	+266.7%

Postoje mnoge različite definicije društvene mreže, no generalizirano gledajući možemo reći da društvene mreže predstavljaju web-usluge koje stavljaju naglasak na ulogu zajedničkog interesa koji međusobno povezuje korisnike društvene mreže. Takve web-usluge omogućuju korisnicima raznovrsne načine interakcije s okruženjem pri čemu svaki korisnik posjeduje vlastiti profil [1].

Detaljni profili korisnika na društvenim mrežama, gdje se bilježi svaka aktivnost korisnika, omogućuju provođenje raznovrsnih analiza. Takva vrsta podataka, gdje korisnici samostalno kreiraju svoje profile i navode osobne podatke, omogućuje velike mogućnosti identifikacije preferencija i navika korisnika te postoji opravdani strah od narušavanja privatnosti i zloupotreba takvih podataka.

Društvenim mrežama danas se koriste stotine milijuna ljudi, a osim onih neutralnih (generalne uporabe) postoje i usluge s određenom namjenom glede sadržajnog fokusa i profila korisnika. Tako svrstavamo društvene mreže Facebook, Twitter, MySpace i sl. u skupinu generalne upotrebe, dok društvenu mrežu LinkedIn svrstavamo u skupinu poslovnih društvenih mreža.

Formalni opis društvene mreže temelji se na teoriji grafova. Dva najčešće upotrebljavana formalizma za opis društvene mreže su *graf* i *matrica*. Graf je ujedno i najčešći oblik vizualizacije strukture društvene mreže gdje su čvorovi profili korisnika, a grane (veze) između njih predstavljaju povezanost. Općenito možemo reći da je graf matematička struktura koja se koristi za opisivanje relacija između dvaju objekata iz određene kolekcije. Matrica u kojoj je zapisana struktura društvene mreže zovemo *matrica susjedstva*. Sami redci i stupci matrice predstavljaju entitete društvene mreže, odnosno skup čvorova grafa, dok svaki unos u pojedinom retku i stupcu označava jesu li entitete (čvorovi) koje predstavljaju odgovarajući redak i stupac susjedni ili nisu.

Definicija: Graf G je uređeni par $G = (V, E)$, gdje je V neprazan skup vrhova, a E je skup bridova. Svaki brid $e \in E$ spaja dva vrha $u, v \in V$ koji se zovu krajevi od e [24].

Definicija: Matrica je pravokutna shema brojeva oblika

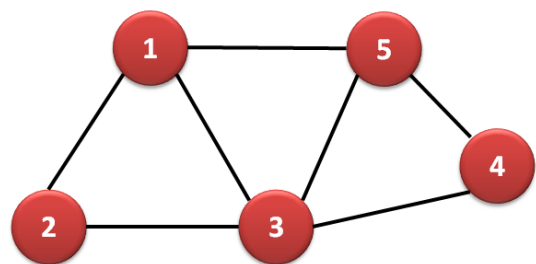
$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}, \quad (13)$$

gdje su $a_{ij} \in R, i = 1, \dots, m; j = 1, \dots, n$. Brojeve a_{ij} nazivamo elementi matrice, gdje indeks i predstavlja broj redaka, a indeks j broj stupaca u kojem se element nalazi. Matrice označavamo velikim slovima. Formula (13) je *opći*

oblik matrice. Ona ima m redaka i n stupaca. Kažemo da je *tipa* (m, n) (ili *reda* (m, n)) [25].

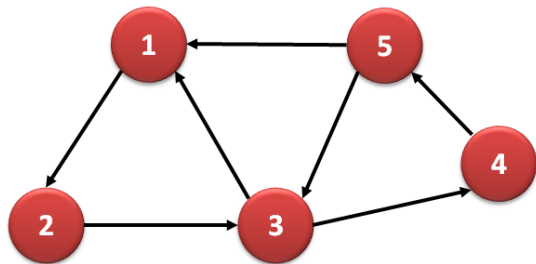
Veze između čvorova (profila korisnika) mogu biti *neusmjerene* i *usmjerene*. Neusmjerena veza predstavlja simetričnu povezanost, npr. ako je čvor X prijatelj s čvorom Y tada vrijedi i da je čvor Y prijatelj s čvorom X . Kod usmjerene veze povezanost je jednosmjerna, npr. ako se čvoru X sviđa sadržaj čvora Y to ujedno ne znači da se čvoru Y sviđa sadržaj čvora X . Graf kojemu su svi bridovi usmjereni zovemo usmjereni graf, u suprotnom, zovemo ga neusmjereni.

Definicija: Neusmjereni graf (Slika 2) je uređeni par skupa vrhova i skupa bridova (V, E) pri čemu su svakom bridu pridružena dva vrha koje on povezuje, pri čemu ta dva vrha mogu biti jedan te isti [26]. ($V = \{1, 2, 3, 4, 5\}$, $E = \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{1, 3\}, \{4, 5\}, \{1, 5\}, \{3, 4\}\}$).



Slika 2. Neusmjereni graf

Definicija: Usmjereni graf (Slika 3) je uređen par $G = (V, A)$ skupa vrhova V i relacije A na V koja ima svojstvo da $(v, w) \in A \implies (w, v) \notin A$. Elemente od V nazivamo vrhovima ili čvorovima grafa [26]. ($V = \{1, 2, 3, 4, 5\}$, $E = \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{3, 1\}, \{4, 5\}, \{5, 1\}, \{3, 4\}\}$).

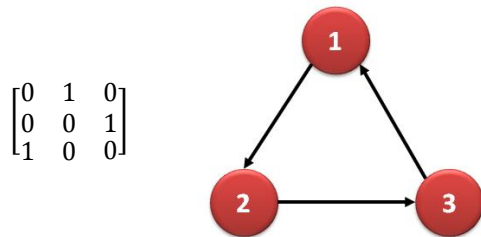


Slika 3. Usmjereni graf

Graf se naziva jednostavnim ako je neusmjeren, nema petlji i između bilo koja dva vrha nema više od jednog brida. U jednostavnom grafu svaki se brid može identificirati s parom različitih vrhova. Brid povezuje dva vrha.

Definicija: Jednostavni graf G sastoji se od nepraznog konačnog skupa $V(G)$, čije elemente zovemo vrhovi (čvorovi) grafa G i konačnog skupa $E(G)$ različitih dvočlanih podskupova skupa $V(G)$ koje zovemo bridovi. Skup $V(G)$ zovemo skup vrhova, a skup $E(G)$ zovemo skup bridova [24].

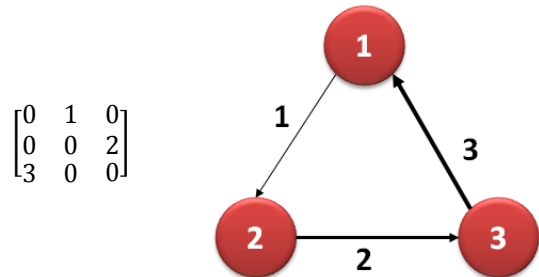
Mrežu možemo definirati grafom (N, g) koji sadrži skup čvorova $N = \{1, \dots, n\}$ i $n \times n$ matricu $G = [g_{ij}]$, gdje g_{ij} predstavlja težinu veze između čvorova i i j . Ukoliko su težine binarne tj. $g_{ij} \in \{0, 1\}$ tada naznačuju postojanje odnosno nepostojanje veze između čvorova i i j . Graf je usmjeren ako vrijedi $g_{ij} \neq g_{ji}$ [27].



Slika 4. Primjer usmjerene mreže

težinski graf i potreban je, primjerice, kod društvene mreže u kojoj sudionici mogu razlikovati prijatelje po interakciji, sličnim interesima i sl. [1].

Definicija: Težinski graf (Slika 6) G^α je graf G čijim su bridovima pridruženi neki realni brojevi, tj. postoji *težinska funkcija* $\alpha : E(G) \rightarrow R$ pri čemu broj $\alpha(e)$ zovemo *težinom brida* $e \in E(G)$ [28].



Slika 6. Težinski graf

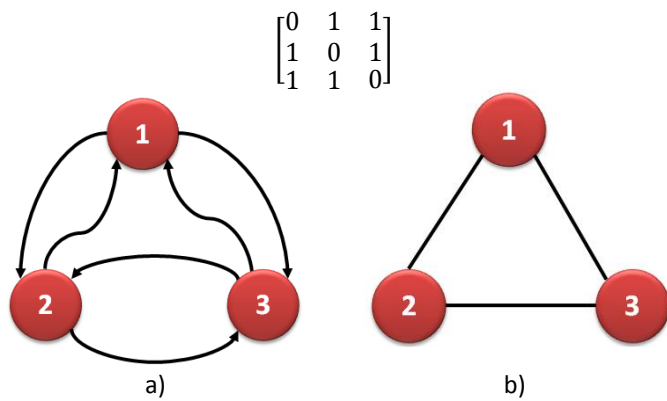
IV. DRUŠTVENI PREPORUČITELJ

Tradicionalni preporučiteljski sustavi kod računanja preporuka/predikcije ne uzimaju eksplicitnu društvenu povezanost među korisnicima. Međutim važnost društvenog utjecaja je odavno prepoznata i priznata u području marketinga [29][30][31][32][33]. Kada potrošač želi kupiti neki proizvod s kojim nije upoznat i nema previše informacija o kvaliteti proizvoda tada se najčešće intuitivno savjetuje s prijateljima, u koje ima povjerenja te koji su već isprobali proizvod ili imaju više informacija o proizvodu. Kada potrošač dobije preporuku od prijatelja za neki određeni proizvod tada takva preporuka ima tendenciju da se prihvati jer je takva vrsta informacije pouzdana. Postoje mnoge marketinške strategije koje su iskoristile ovaj aspekt ljudske prirode čime su postigle veliki uspjeh [35]. Primjer takve uspješne marketinške kampanje je promocija besplatnog elektroničnog sandučića Hotmail¹⁹. Marketinška strategija je podrazumijevala automatsko dodavanje promotivne poruke za Hotmail na svaki poslani mail. Kao rezultat ove kampanje je porast broja korisnika Hotmaila od nula do osamnaest milijuna korisnika za osamnaest mjeseci uz budžet za oglašavanje od pola milijuna dolara [35]. Ovakav način oglašavanja je nadmašio mnoge konvencionalne marketinške strategije [34]. Dakle pokazano je da društveni utjecaj ima ključnu ulogu kod potrošača kada donosi odluku o korištenju neke usluge ili kupnje nekog proizvoda.

A. Izvori informacija o društvenom kontekstu

Nedavna pojava i rast popularnosti društvenih mreža te omogućavanje aplikacijskog programskog sučelja (engl. *application programming interface*) za dohvaćanje podataka s društvenih mreža omogućeno je istraživanje uloge društvenog utjecaja u području preporučiteljskih sustava. Uz

¹⁹ www.hotmail.com



Slika 5. Primjer usmjerene (a) i neusmjerene mreže (b)

Slika (Slika 4) prikazuje mrežu definiranu matricom 3×3 i usmjerenim grafom s tri čvorova. Vrijednosti u matrici prikazuju vrijednosti težina grana. Redovi u matrici označavaju početak usmjerene veze, dok stupci označuju kraj usmjerene veze za pojedini čvor. Tako da imamo u prvom retku i drugom stupcu vrijednost 1 koja označuje usmjerenu vezu iz čvora 1 prema čvoru 2. Na slici (Slika 5) je također prikazana mreža definirana matricom 3×3 i s dva grafa. Prvi graf je usmjereni, dok je drugi graf neusmjereni. Ta dva grafa su identična, ako su vrijednosti težina veza usmjerenih iz jednog čvora u drugi i iz drugog čvora u prvi jednake.

Za razliku od neusmjerenih i usmjerenih grafova, pogodnih za modeliranje društvenih mreža u kojima određena veza između entiteta ili postoji ili ne postoji, za modeliranje društvenih mreža često su potrebni grafovi koji mogu prikazati težinu (intenzitet, jačinu) veze između pojedinog para entiteta. Graf kojim se želi modelirati društvena mreža u kojoj su veze između entiteta težinski označene i predstavljaju intenzitet određene veze mora moći prikazati takve veze svojim skupom grana. Ovakav tip grafa naziva se

rast popularnosti Weba 2.0 pojavile su se mnoge društvene mreže kao što su Facebook²⁰, LinkedIn²¹, Twitter²², Myspace²³ i dr. S obzirom na vrstu društvene mreže postoje različite veze između korisnika takvih društvenih mreža. Vidljive i dostupne veze između korisnika mogu biti jednosmjerne, dvosmjerne i udaljene (razne razine prijateljstva).

Važan izvor podataka i društvenog konteksta za sustava za preporuku filmova su razne specijalizirane web-stranice, a neke od najpopularnijih su IMDb i TMDb²⁴. Također, neizostavni izvori društvenog konteksta za filmove su društvena mreža Facebook i popularna internetska usluga za razmjenu video sadržaja YouTube.

IMDb (engl. *The Internet Movie Database*) je najpopularnija baza filmova na svijetu te predstavlja autoritativni izvor informacija o filmovima, televizijskim emisijama i slavnim osobama. Baza prikuplja u prosjeku oko 100 milijuna jedinstvenih posjeta mjesečno, a omogućuje pretraživanje podataka o više od 1.8 milijuna filmova, televizijskih emisija i zabavnog programa. Baza sadrži više od 4 milijuna zapisa o slavnim osobama i osoblju u filmskoj industriji.

Baza filmova TMDb započela je s radom u jesen 2008. godine. Prvotno je bila zamišljena kao stranica koja će sadržavati slike i postere visoke rezolucije u komprimiranom zip-formatu za popularni projekt XBMC. Projekt XBMC je nagrađivani besplatan program otvorenog koda (engl. *open source*) i mjesto gdje se može pronaći zabavna vrsta digitalnog sadržaja. Nakon što je pokrenuta kao jednostavna web-stranica koja je poveznicama omogućavala dohvaćanje komprimiranih datoteka s vremenom je prerasla u jednu od najvećih baza za filmove na svijetu.

U radu [36] baza filmova TMDb predstavlja početak ciklusa prikupljanja podataka o filmovima te ujedno dohvaćeni profili filmova iz TMDb-a predstavljaju početnu točku za pretraživanje dodatnih podataka o filmovima:

- u bazi filmova IMDb;
- popularnoj usluzi za razmjenu video sadržaja YouTube; i
- društvenoj mreži Facebook.

Dohvaćeni podaci o filmovima s IMDb-a su sve dostupne ocjene po demografskim kategorijama, s YouTube se dohvaćaju broj pregleda trailer videa, broj korisnika koji su ocijenili video i prosječna ocjena videa. Posljednji korak u procesu prikupljanja informacija o filmovima koristi se

društvena mreža Facebook, gdje se pronalaze službene Facebook stranice filmova te se u bazu sprema podatak o broju korisnika koji su se izjasnili da im se film sviđa. Podaci o korisnicima prikupljaju se posredstvom društvene mreže Facebook. Korištenjem društvenog grafa omogućen je pristup profilima korisnicima iz kojih su dostupne osobne informacije o korisnicima i podaci o sadržajima koje korisnici vole [36].

B. Metode društvenog preporučitelja

U nastavku su opisane metode društvenog preporučitelja koje su obrađene u istraživačkim radovima.

Autori u radu [37] istražuju pojam povjerenja na društvenim mrežama te na temelju prikupljenih podataka s društvene mreže Facebook generiraju listu najbližih prijatelja. Lista najbližih prijatelja je dobivena iz implicitne društvene mreže generirane na temelju međusobne interakcije korisnika na društvenoj mreži Facebook. Ideja generiranja implicitne društvene mreže obuhvaća transformaciju diskretne veze (veza postoji ili ne tj. 1 ili 0, Slika 5) u kontinuirane težinske veze, gdje težine označuju povjerenje između korisnicima (Slika 6). Provedeno je istraživanje na 150 korisnika koji su koristili razvijenu Facebook aplikaciju Closest Friends koja je nudila korisnicima različite liste najbližih prijatelja izračunate različitim težinama razmatranih podataka (Facebook fotografije korisnika, interakcija putem korisničkog Facebook zida, korisničkog Facebook pretinca).

Nadalje u radu [38] autori istražuju primjenu implicitnih društvenih mreža u multinacionalnoj kompaniji. Za potrebe izgradnje implicitne društvene mreže razvijeni su sustavi za analizu elektroničke pošte, telefonskih poziva i stvarnovremenskih poruka. Također je razvijena aplikacija koja omogućuje različite analize društvene mreže pomoću koje se detektiraju važni čvorovi i grupe unutar takve mreže.

U radu [36] je razvijen studijski slučaj društvenog preporučitelja koji se sastoji od tri glavna dijela izvori informacija, sustav za preporuku i korisničke aplikacije. Komponenta izvori informacija predstavlja proces prikupljanja informacija o filmovima i korisnicima. Središnji dio društvenog preporučitelja je sustav za preporuku koji obuhvaća proces obrade prikupljenih informacija i preporuku filmova. Posljednja komponenta društvenog preporučitelja je korisnička aplikacija koja omogućuje krajnjim korisnicima pregled preporučenih filmova i liste najbolje rangiranih prijatelja.

²⁰ www.facebook.com

²¹ www.linkedin.com

²² www.twitter.com

²³ myspace.com

²⁴ <http://www.themoviedb.org/> (The Movie Database)

Na temelju prikupljenih podataka o korisnicima i njihove međusobne interakcije na društvenoj mreži Facebook izračunava se [36]:

- koje žanrove korisnik najviše voli;
- izračunava se indeks interakcije na društvenoj mreži Facebook između korisnika;
- izračunava se sličnost između korisnika koji koriste razvijenu Facebook aplikaciju *FERmovies* na temelju:
 - filmova koje vole;
 - sadržaja za koje su se izjasnili da vole;
 - Facebook grupa u koje su korisnici pridruženi; i
 - interakcije između prijatelja;

Nakon izračunatih indeksa interakcije i indeksa sličnosti između korisnika aplikacije *FERmovies* korisniku se preporučuju filmovi, iz najbolje rangiranih žanrova, za koje su se tri njegova najbližnja Facebook prijatelja izjasnila da im se sviđaju, a da se korisnik za te filmove nije već izjasnio da mu se sviđaju. Također se korisniku preporučuju najpopularniji filmovi po žanrovima koji su najbolje rangirani na temelju [36]:

- ukupnog broja pregledanih *trailer* filmova na YouTube;
- ukupnog broja korisnika koji su se izjasnili da im se film sviđa na bilo kojoj od analiziranih poveznica;
- ukupnog broja korisnika koji su ocijenili film;
- izračunatoj prosječnoj ocjeni filma na temelju prikupljenih svih ocjena filma.

U radu [35] su razmatrane eksplicitne društvene veze u preporučiteljskim sustavima, npr. kako su korisničke preferencije i ocjene u korelaciji s preferencijama i ocjenama prijatelja. U radu je dizajniran algoritam za izračun preporuka koji se temelji na korisničkim preferencijama i mišljenjima njihovih prijatelja. Podaci nad kojim je rađena analiza su dohvaćeni s društvene mreže Yelp²⁵. Rezultati eksperimenta su pokazali da je točnost predviđanja poboljšana za 17.8% u odnosu na tradicionalni preporučitelj temeljene na suradnji korištenjem metode MAE (engl. *mean absolute error*).

Članak [39] obuhvaća područje pretraživanja i proširenih upita (engl. *query expansion*, QE). Prošireni upiti omogućuju korisnicima proširenje pretraživane domene dopunjivanjem njihovih početnih upita dodatnim uvjetima i frazama [40].

²⁵ <http://www.yelp.com/> - je američka web-stranica koja omogućava ocjenjivanje i recenzije raznog sadržaja sa značajkama društvene mreže

Primjer:

Početni upit	Kategorizacija, oznake	Prošireni upit
amazon	e-trgovina, kupovina	buy AND (books OR book) AND amazon
amazon	priroda	(rivers OR river) AND amazon

Tradicionalne tehnike proširenih upita QE temelje se na računanju dvodimenzionalnih korelacijskih matrica [41]. U radu je na tradicionalnu dvodimenzionalnu matricu nadodana treća dimenzija koja predstavlja društvenu komponentu. Treća dimenzija su semantičke klase (odnosno, kategorije koje obuhvaćaju uvjete određenog semantičkog značenja) generirane na temelju usluga društvenog bookmarkinga kao što su *Delicious*²⁶, *StumbleUpon*²⁷ i *Digg*²⁸. Rezultati rada potvrđuju korelaciju između generiranih (preporučeni) rezultata i interesa korisnika te je ujedno time pokazana važnost kategorizacije obilježja (engl. *tag*) u semantičke klase.

Autori u članak [42] navode da se preciznost preporuka može poboljšati implementacijom veza povjerenja (engl. *trust relationships*) između korisnika preuzimanjem tih veza s društvenih mreža. Preciznost preporuka se u ovome slučaju mjeri metodom korijena srednje kvadratne pogreške (engl. *root mean square error*, RMSE). U radu je provedeno istraživanje kako poboljšati preciznost top-k preporuka korištenjem društvenih mreža. Također se predlaže da top-k preporuke temeljene na metodi najbližih susjeda (engl. *nearest neighbor*) kombiniraju „najbliže“ korisnike iz mreže povjerenja i „najbliže“ korisnike dobivene klasteriranjem korisnika korištenjem Pearsonovog koeficijenta korelacije. Rezultati eksperimenata na dva dostupna skupa podataka (*Epinions*²⁹ i *Flixster*³⁰) pokazuju da društvene mreže mogu značajno poboljšati top-k preporuke (engl. *top-k hit ratio*, odnos broja zanimljivih i ne zanimljivih preporuka). Ovaj zaključak naročito vrijedi za nove korisnike kojima se javlja problem hladnog starta. U radu je također uočeno da kombinacija povratne veze (npr. ocjene sadržaja) i društvenih podataka koji pozitivno utječu na minimizaciju RMSE negativno utječu na maksimizaciju odnos broja zanimljivih i ne zanimljivih preporuka.

Autori u radu [43] istražuju kako društvena povezanost (društveni graf) utječe na sličnost između ukusa korisnika. Analiza je provedena na skupu podataka švedske internetske

²⁶ <https://delicious.com/>

²⁷ <http://www.stumbleupon.com/>

²⁸ <http://digg.com/>

²⁹ <http://www.epinions.com/> - web-stranica koja omogućuje ocjenjivanje i izražavanje mišljenja korisnicima za razne predmete i usluge

³⁰ <http://www.flixster.com/> - web-stranica koja omogućuje korisnicima ocjenjivanje filmova

stranice za preporuku filmova Filmtipset³¹ na više od 80 000 korisnika. Filmtipset omogućuje asimetrična i simetrična prijateljstva. Za izračunavanje sličnosti između korisnika korišten je koeficijent Jaccardove sličnosti:

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

Dok za izračun ukupne sličnosti cijelog razmatranog društvenog grafa korištena formula:

$$\bar{J} = \frac{2}{n(n-1)} \sum_{i < j} \frac{|A_i \cap B_j|}{|A_i \cup B_j|}$$

Rezultati istraživanja pokazuju da reducirani društveni grafovi po određenim pravilima (izbacivanje korisnika koji nemaju barem 5 prijatelja grupirati u jedan društveni graf, grupirati korisnike koji nemaju uopće prijatelje i sl.), povećavaju ukupno sličnost između preostalih korisnika te time omogućuju poboljšanje kvalitete preporuke.

V. ZAKLJUČAK

U radu su opisane dvije glavne skupine preporučiteljskih sustava:

- preporučitelji temeljeni na sadržaju i;
- preporučitelji temeljeni na suradnji.

Prikazani su osnovni mehanizmi preporučiteljskih sustava te su navedene njihove prednosti i nedostaci. U vidu poboljšanja preporučiteljskih sustava uveden je koncept društvenog konteksta. Društvene mreže danas kao jedna od najpopularnijih usluga na internetskoj infrastrukturi koja generira veliku količinu podataka predstavlja bogat izvor društvenog konteksta. Društveni kontekst predstavljaju međusobni odnos i interakcije korisnika na društvenim mrežama te prikazuju preferencije korisnika za određenu vrstu sadržaja. Stoga, društvena preporuka sadržaja predstavlja personaliziranu informaciju jer sadržaj koji se preporučuje takvim inteligentnim sustavom preporučuje sadržaje na temelju ponašanja samoga korisnika i njegove interakcije s okruženjem.

Na temelju razmatranja zaključujemo da integracija društvenog konteksta u preporučiteljske sustave ima pozitivne efekte na njihovo funkcioniranje [35]:

- Prvo, u vidu poboljšanja točnosti predviđanja jer dobivene dodatne informacije o korisnicima i njihovim prijateljima olakšavaju razumijevanje ponašanja korisnika i njihovih ocjena. Stoga možemo modelirati i interpretirati korisničke preferencije mnogo preciznije i time poboljšati točnost predviđanja.

- Drugo, s informacijom o prijateljima iz društvenih mreža više nije nužno pronaći slične korisnike mjerenjem sličnosti jer činjenica da su dva korisnika prijatelji već ukazuje samo po sebi da imaju nešto zajedničko. Ovom pretpostavkom možemo ublažiti problem raznolikosti preporuka (engl. *sparsity problem*).
- Treće, izbjegavamo problem hladnog starta ukoliko koristimo preporučitelj temeljen na suradnji, iako korisnik nema nikakve aktivnosti u prošlosti, jer mu možemo izračunati preporuke temeljene na preferencijama njegovih prijatelja.

Sva ova promatranja su motivacija za integraciju društvenih podataka u preporučiteljske sustave te time želimo iskoristiti prednosti novo dostupnih podataka s društvenih mreža.

Da bi iskoristili sve navedene pozitivne efekte društvenog konteksta u preporučiteljskim sustavima susrećemo se s izazovima u istraživanju:

- novo dostupni podaci o korisnicima i njihovim ponašanjima na Internetu;
- problemi velike količine podataka (engl. *big data*) te pojava i korištenje novih tehnologija za njihovu obradu (ne-SQL baze podataka, engl. *No-SQL*);
- primjena različitih matematičkih i statističkih modela za izračunavanje preporuka.

Nadalje, poduzeća su također prepoznala veliku vrijednost društvenog konteksta za unapređenje poslovanja. Tako da je unapređenje poslovanja za poduzeća posebno izraženo u vidu poboljšanja kvalitete usluge jer su usluge bolje personalizirane. Također postoje velike mogućnosti primjene u području marketinga kod razvoja proizvoda i usluga te interakcije s potrošačima. Pojavom društvenih mreža pojavile su se nove i/ili unaprijeđene mogućnosti za primjene u marketingu: nova vrsta preporučenoga sadržaja („od usta do usta“), istraživanje tržišta, kreiranje i provjera ideja, razvoj novih proizvoda u suradnji s potrošačima, unapređenje odnosa s kupcima, najnaprednije načine promocije (npr. personalizirani oglasi) i ostalo.

Na temelju provedenog razmatranja možemo reći da se koncept društvenog konteksta integrira u različite sustave s ciljem razvoja ili unapređenja bilo usluga, proizvoda ili samog poslovanja.

³¹ <http://nyheter24.se/filmtipset/>

LITERATURA

- [1] Podobnik V., Višeagentski sustav za pružanje telekomunikacijskih usluga zasnovan na profilima korisnika, doktorska disertacija, Zagreb, 2010.
- [2] Štriga D., Analiza interakcije s markama putem društvenih mreža, diplomski rad, Sveučilište u Zagrebu, Ekonomski fakultet, Zagreb, 2012.
- [3] Internet Marketing vs. Traditional Marketing, <http://www.netlz.com/seoblog/2011/11/16/internet-marketing-vs-traditional-marketing/> (rujan 2013.)
- [4] The Advantages Of Internet Marketing Vs. Traditional Marketing, <http://www.articledashboard.com/Article/The-Advantages-of-Internet-Marketing-vs-Traditional-Marketing/652141>, (rujan 2013.)
- [5] Types of Marketing: Traditional & Internet, <http://homebusiness.about.com/od/marketingadvertising/a/Types-Of-MarketingTraditional-And-Internet.htm>, (rujan 2013.)
- [6] Online & Mobile Advertising Methods Outperform Traditional Advertising, <http://www.jvtimes.com/4/online-mobile-advertising-methods-outperform-traditional-advertising/>, (rujan 2013.)
- [7] Pazos Arias, José J., Fernández Vilas, Ana, Díaz Redondo, Rebeca P., Recommender Systems for the Social Web, Intelligent Systems Reference Library, Vol. 32, Springer, 2012.
- [8] Groh G., Lehmann A., Reimers J., Friess M.R., Schwarz L., Detecting Social Situations from Interaction Geometry, SOCIALCOM '10 Proceedings of the 2010 IEEE Second International Conference on Social Computing, Pages 1-8 , IEEE Computer Society Washington, DC, USA 2010
- [9] Pazzani M. J., Billsus D., Content-based recommendation systems, The adaptive web, Pages 325-341, Springer-Verlag Berlin, Heidelberg 2007
- [10] Ricci F.; Rokach L.; Shapira B.; Kantor P.B., Recommender Systems Handbook, Springer 2011
- [11] Salton G., Automatic text processing: the transformation, analysis, and retrieval of information by computer, Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA 1989
- [12] Jannach D., Zanker M., Felfernig A., Friedrich G., Recommender Systems: An Introduction, Cambridge University Press; 1 edition, 2010
- [13] Pauše Ž., Uvod u matematičku statistiku, Zagreb: Školska knjiga, 1993
- [14] Herlocker J. L., Konstan J. A., et al., An Algorithmic Framework for Performing Collaborative Filtering, Proceedings of the 22nd Annual International ACM SIGIR Conference, ACM Press, 1999
- [15] Breese J. S., Heckerman D., Kadie C. M., Empirical analysis of predictive algorithms for collaborative filtering, Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann, 1998, pp. 43–52.
- [16] Anand S. S., Mobasher B., Intelligent techniques for web personalization, Lecture Notes in Computer Science, vol. 3169, Springer, Acapulco, Mexico, 2005, pp. 1–36.
- [17] Herlocker J., Konstan J. A., Riedl J., An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms, Information Retrieval 5 (2002), no. 4, 287–310
- [18] Sarwar B., Karypis G., Konstan J., Riedl J., Analysis of recommendation algorithms for e-commerce, Proceedings of the 2nd ACM Conference on Electronic Commerce, ACM, 2000, pp. 158–167.
- [19] Sarwar B., Karypis G., Konstan J., Riedl J., Item-based collaborative filtering recommendation algorithms, Proceedings of the 10th International Conference on World Wide Web (WWW '01) (Hong Kong), ACM, 2001, pp. 285–295.
- [20] Goldsmith B., <http://www.reuters.com/article/2008/09/16/us-internet-book-life-idUSSP31943720080916> (rujan 2013)
- [21] Qualman E., Social Media Video 2013, <http://www.socialnomics.net/2012/11/07/social-media-video-2013/> (rujan 2013)
- [22] Fox Z., This Is How Much Time You Spend on Facebook, Twitter, Tumblr, <http://mashable.com/2012/11/28/social-media-time/> (rujan 2013)
- [23] Thorhauge S., How People Spend Their Time Online, <http://www.mindjumpers.com/blog/2012/05/time-spend-online/> (rujan 2013)
- [24] Pavčević M. O., Uvod u teoriju grafova, Element, Zagreb, 2006
- [25] Pašić M., Matematička analiza 1, FER, 2004
- [26] Čaklović L., Matematičko modeliranje: Osnove Teorije grafova
- [27] Vega-Redondo F., Complex Social Networks, Cambridge University Press, 2007
- [28] Majstorović S., Težinski grafovi. Problem najkraćeg puta. Dijkstrinov algoritam, Osijek, 2012
- [29] Subramani M. R., Rajagopalan B., Knowledge-sharing and influence in online social networks via viral marketing. Communications of the ACM, 46(12):300–307, 2003.
- [30] Taskar B., Abbeel P., Koller D., Discriminative probabilistic models for relational data. In Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence (UAI), Edmonton, AB, pp. 485–492, 2002.

- [31] Ungar L.H., Foster D.P., Clustering methods for collaborative filtering. In Workshop on Recommender Systems at the 15th National Conference on Artificial Intelligence, Madison, WI, 1998.
- [32] Wang J., Vires A.P., Reinders M.J.T., Unifying user-based and item-based collaborative filtering approaches by similarity fusion. In Proceedings of the 29th Annual International ACM SIGIR Conference on Research & Development on Information Retrieval (SIGIR06), August 6–11, 2006.
- [33] Yang S., Allenby G.M., Modeling interdependent consumer preferences. *Journal of Marketing Research*, 40:282–294, 2003.
- [34] Jurvetson S., What exactly is viral marketing? *Red Herring*, 78:110–112, 2000.
- [35] Chu W.W., He J., A social network-based recommender system, doctoral dissertation, University of California at Los Angeles Los Angeles, CA, USA 2010
- [36] Štriga D., Preporučivanje sadržaja zasnovano na društvenoj umreženosti korisnika, diplomski rad, Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva, lipanj 2011.
- [37] Podobnik V., Štriga D., Jandras A., Lovrek I., How to Calculate Trust between Social Network Users? Proceedings of the 20th International Conference on Software, Telecommunications and Computer Networks (SoftCOM 2012). Split, Croatia: IEEE, 2012.
- [38] Humski L., Štriga D., Podobnik V., Vrdoljak B., Banek M., Skocir Z., Lovrek I., Building Implicit Corporate Social Networks: the Case of a Multinational Company, Proceedings of the 12th International Conference on Telecommunications, Zagreb, Croatia: IEEE, 2013. 31-38.
- [39] Biancalana C., Gasparetti F., Micarelli A., Miola A., Sansonetti G., Folksonomy-based adaptive query expansion, in Proc. UMAP Workshops, 2012.
- [40] Bai J., Song D., Bruza P., Nie J.-Y., Cao G., Query expansion using term relationships in language models for information retrieval. In Proceedings of the 14th ACM International Conference on Information and Knowledge Management, pages 688–695, 2005.
- [41] Biancalana C., Lapolla A., Micarelli A., Personalized web search using correlation matrix for query expansion. *WEBIST*, pages 186–198, 2008.
- [42] Yang X., Steck H., On Top-k Recommendation using Social Networks Categories and Subject Descriptors, *RecSys'12*, September 9–13, 2012, Dublin, Ireland, UK
- [43] Said A., De Luca E. W., Albayrak S., How social relationships affect user similarities, Workshop