

UDŽBENIK ELEKTROTEHNIČKOG FAKULTETA U BEOGRADU

Milan Bjelica

# PERSONALIZOVANE APLIKACIJE

TEORIJA I PRAKSA

Beograd, 2018.

dr Milan Bjelica,  
Univerzitet u Beogradu – Elektrotehnički fakultet  
e-mail: milan@etf.rs

PERSONALIZOVANE APLIKACIJE: TEORIJA I PRAKSA  
*elektronski udžbenik*

Recenzenti:

prof. dr Aleksandra Smiljanić,  
doc. dr Mirjana Simić-Pejović

Nastavno-naučno veće Elektrotehničkog fakulteta odobrilo je objavljivanje ovoga udžbenika odlukom broj 1314/4 od 12.7.2016. godine.

Izdaje i štampa:  
Elektrotehnički fakultet  
Beograd, 2018.

Tiraž: 50 primeraka

ISBN: 978-86-7225-059-6



Delo je licencirano pod uslovima licence  
Creative Commons  
Autorstvo – Deliti pod istim uslovima 4.0

Tekst ove knjige složen je u programskom paketu L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub>.

---

CIP - Каталогизација у публикацији - Народна библиотека Србије, Београд  
004.42:004.9(075.8)(0.034.2)

БЈЕЛИЦА, Милан, 1977-

Personalizovane aplikacije [Elektronski izvor] : teorija i praksa :  
[elektronski udžbenik] / Milan Bjelica. - Beograd : Univerzitet,  
Elektrotehnički fakultet, 2018 (Beograd : Univerzitet, Elektrotehnički  
fakultet). - 1 elektronski optički disk (CD-ROM) ; 12 cm. - (Udžbenik  
Elektrotehničkog fakulteta u Beogradu)

Sistemski zahtevи: Nisu navedeni. - Nasl. sa naslovne strane dokumenta. -  
Tiraž 50. - Sadrži bibliografiju.

ISBN 978-86-7225-059-6

a) Апликативни софтвер - Пројектовање - Персонализација  
COBISS.SR-ID 258998284

---

# Sadržaj

<b>1</b>	<b>Uvod u personalizaciju</b>	<b>1</b>
<b>2</b>	<b>Pronalaženje informacija</b>	<b>5</b>
2.1	Uvod . . . . .	5
2.2	Definicija . . . . .	5
2.3	Osnovni elementi . . . . .	6
2.4	Tehnike pronalaženja informacija . . . . .	8
2.5	Osnovni probabilistički model . . . . .	11
2.6	Model Bayesovih mreža . . . . .	14
2.7	Fazi model . . . . .	19
2.8	Performanse sistema za pronalaženje informacija . . . . .	20
2.9	Pitanja i zadaci . . . . .	24
<b>3</b>	<b>Model vektorskog prostora</b>	<b>27</b>
3.1	Uvod . . . . .	27
3.2	Opis . . . . .	27
3.3	Skalarni proizvod vektora . . . . .	29
3.4	Primena VSM za pronalaženje informacija . . . . .	30
3.5	Modeliranje multimedijalnog dokumenta . . . . .	34
3.6	Relevantnost dokumenta za upit . . . . .	35
3.6.1	Mere razdaljine . . . . .	35

3.6.2	Mere sličnosti . . . . .	38
3.7	Pitanja i zadaci . . . . .	40
<b>4</b>	<b>Modeliranje korisnika</b>	<b>43</b>
4.1	Uvod . . . . .	43
4.2	Sistemi za modeliranje korisnika . . . . .	43
4.3	Reakcija po značaju . . . . .	48
4.4	Problem zaštite privatnosti . . . . .	49
4.5	Modeliranje korisnika u vektorskim prostorima . . . . .	51
4.6	Pitanja i zadaci . . . . .	54
<b>5</b>	<b>Personalizacija u praksi</b>	<b>55</b>
5.1	Uvod . . . . .	55
5.2	Preporučivači sadržaja . . . . .	56
5.2.1	Sistemi zasnovani na opisu sadržaja . . . . .	57
5.2.2	Sistemi zasnovani na kolaborativnom filtriranju . . . . .	59
5.2.3	Hibridni sistemi . . . . .	61
5.2.4	Sistemi zasnovani na prepoznavanju konteksta . . . . .	62
5.3	Pretraživači interneta . . . . .	63
5.4	Aplikacije za mobilne terminale . . . . .	67
5.5	Pitanja i zadaci . . . . .	69
<b>Dodatak:</b>	<b>Personalizovani programske vodiči</b>	<b>73</b>
<b>Literatura</b>		<b>79</b>

# Poglavlje 1

## Uvod u personalizaciju

Predmet razmatranja ove knjige su personalizovane aplikacije, a posmatrano šire i telekomunikacioni servisi, koje se odlikuju mogućnošću prilagođavanja svojih karakteristika procenjenim željama, potrebama ili interesovanjima konkretnog korisnika.

Potreba za personalizovanim aplikacijama jasno je uočljiva u današnjem dinamičkom okruženju multiservisnih telekomunikacionih mreža. Operatori mreža, odnosno nudioci servisa i sadržaja, moraju iznaći strategije za očuvanje postojećih i pridobijanje novih korisnika; pokazaćemo da personalizacija može da bude rešenje ovog problema. Naime, u uslovima liberalizacije tržišta, zadovoljstvo korisnika pruženim servisom postaje bitan faktor koji ga opredeljuje za izbor operatora mreže ili nudioca servisa i sadržaja, što Međunarodna unija za telekomunikacije jasno naglašava u preporuci G.1000 [ITU01]. Zadovoljan korisnik bez sumnje intenzivnije i učestalije koristi servis, pa time donosi i veći profit svom operatoru.

Ilustrativan primer za opravdanost uvođenja personalizacije nalazimo kod kablovske televizije. Prelazak na digitalnu distribuciju programa omogućio je nuđenje sadržaja koji po obimu i prirodi ranije nisu bili zamislivi. Uočavamo dva načina na koje se operatori mogu boriti za stare i nove korisnike [Pet06]:

1. Operatori mogu uvećavati fond televizijskih programa i neselektivno ih nuditi svim korisnicima, ili
2. Operatori mogu da procene interesovanja korisnika i da onda iz fonda raspoloživih programa izaberu one za koje prepostavljaju da će interesovati posmatranog korisnika.

Prvu strategiju ćemo nazvati strategijom kvantiteta, a drugu strategijom kvaliteta. Razmotrimo ukratko njihove odlike.

Strategija kvantiteta i dalje se najčešće primenjuje. Motivisana je očekivanjem da će među velikim brojem sadržaja koji se nude svaki korisnik pronaći nešto što ga u datom trenutku interesuje. Pri tome, u najmanju ruku je diskutabilno hoće li prosečan kori-

snik ikada posvetiti ozbiljniju pažnju nekim od stotina raspoloživih programa. Samo upoznavanje s njihovim sadržajem, bilo kroz TV magazine ili „menjanjem“ kanala, oduzelo bi previše vremena; verovatnijom se čini varijanta da će se korisnik zadržavati samo na nekolicini programa koji ga zaista interesuju, po cenu i da propusti zanimljiji sadržaj koji se u isto vreme nudi na nekom od, za njega, nepoznatih kanala. Ovo pre ili kasnije kod korisnika izaziva negativan stav prema nudiocu, jer plaća za pristup velikom broju programa koje efektivno ne koristi. Strategiju kvantiteta stoga bismo mogli nazvati i konfekcijskom strategijom, jer ona ne uzima u obzir želje i očekivanja pojedinačnih korisnika.

Nasuprot tome, u strategiji kvaliteta korisniku se neposredno nudi samo nekolicina programa, ali upravo onih koji bi trebalo da za njega budu visoko vredni. Prednosti ovakvog pristupa očigledno su višestruke: štede se resursi mreže, jer se do korisnika ne prenose svi raspoloživi sadržaji; takođe, subjektivni osećaj zadovoljstva korisnika raste, jer on stiče utisak da nudilac ima samo sadržaje koji ga zanimaju, pa stoga više koristi posmatrani servis. Analogiju s ovom situacijom nalazimo u trgovini, kada je s pojavom samoposluga i približavanjem robe kupcu porasla potrošnja.

Drugi primer za potrebu uvođenja personalizacije pruža nam tzv. paradoks aktivnog korisnika [Car87]. Istraživanja su pokazala da korisnici u opštem slučaju imaju probleme pri korišćenju tehničkih uređaja, npr. telekomunikacionih terminala, pri čemu se njihov prosečan nivo znanja ne menja s razvojem tehnike, već asimptotski teži osrednjem. Tipičan korisnik ne čita uputstvo za upotrebu, već odmah pokušava da koristi uređaj i na taj način obavi posao. Paradoks se ogleda u činjenici da bi korisnici globalno uštedeli vreme kada bi na početku proučili uputstvo i podesili uređaj prema stvarnim potrebama. Projektovanje uređaja za idealnog, racionalnog korisnika, prema tome nema opravdanja u stvarnom svetu; rešenje je u uvođenju adaptivnih korisničkih interfejsa, koji bi estimirali potrebe korisnika i nivo njegovog tehničkog predznanja i na osnovu toga automatski izvršili potrebna podešavanja.

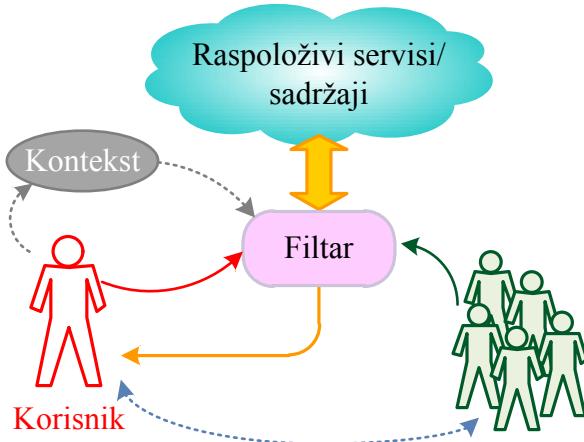
Razmotrimo sada koji bi parametri servisa mogli da budu predmet personalizacije. Posmatrajmo, kao primer, hipotetički servis razmenjivanja poruka. Za njega bismo mogli podešavati:

- Način prezentacije
  - boju pozadine, pismo, jezik rečnika, veličinu, tip i boju slova za tekstualne poruke,
  - paletu boja za grafičke poruke,
  - kvalitet zvuka za audio poruke,
  - rezoluciju za slajdovske i video poruke;
- Standardno zaglavlje i potpis poruke;
- Tehničke parametre
  - rok isporuke,
  - izveštaj o dostavljanju,
  - izveštaj o čitanju,
  - povratnu adresu;
- Cenu

- ograničenje cene poruke,
- slanje po nižoj tarifi,
- prioritetnu isporuku,
- plaćeni odgovor itd.

Vidimo da postoji ogroman prostor za podešavanje vrednosti parametara čak i ovako jednostavnog servisa.

Princip personalizacije ilustrovan je na slici 1.1.



Slika 1.1: Princip personalizacije.

Sistem posmatra neka svojstva i aktivnosti korisnika, npr. njegovu geografsku lokaciju, uzrast, školsku spremu, telefonske brojeve koje je pozivao, internet sajtove koje je posećivao, muziku koju je slušao i na osnovu prikupljenih podataka formuliše kriterijume po kojima iz skupa raspoloživih servisa i sadržaja filtrira one za koje procenjuje da bi trebalo da interesuju toga korisnika. Kriterijum filtriranja može se formulisati i prema aktivnostima drugih korisnika koji su srodni posmatranome, kao i prema kontekstu u kome je posmatrani korisnik pristupao sadržajima.

Vidimo da je za pravilno sprovođenje personalizacije potrebno (1) proceniti potrebe korisnika i (2) izdvojiti ono što ga interesuje. Rešenje ovih problema nalazimo u oblastima modeliranja korisnika i pronalaženja informacija, pa pitanje projektovanja personalizovanih aplikacija sada postaje multidisciplinarni problem.

U nastavku knjige, date su osnove pronalaženja informacija, discipline koja pokazuje na koji se način iz skupa raspoloživih dokumenata efikasno i pouzdano mogu izdvojiti oni koji odgovaraju izrečenim zahtevima korisnika. Ova materija je uglavnom klasična, pa je naročita pažnja posvećena kompromisu između dubine i matematičke korektnosti izlaganja. U trećem poglavlju razmotrena je jedna od metoda pronalaženja informacija, tzv. model vektorskog prostora. Njegova specifičnost ogleda se u tome što se raspoloživi objekti, nezavisno od svoje prirode – bilo da su to tekstualni dokumenti, slike, filmovi,

audio klipovi ili telekomunikacioni servisi u najširem smislu – predstavljaju vektorima svojih odlika. Četvrto poglavlje bavi se pitanjem formalnog matematičkog kvantifikovanja korisnikovih želja, potreba i interesovanja; u njemu su date osnove sistema za modeliranje korisnika. Pokazano je da veliki problem predstavlja zaštita privatnosti podataka o korisnicima, koji se prikupljaju za potrebe personalizacije. Peto poglavlje bavi se primerima personalizovanih aplikacija i to preporučivačima sadržaja, pretraživačima interneta i lokacijskim servisima. Na kraju knjige, u dodatku, detaljnije su razmotreni televizijski programski vodiči.

U tekstu knjige, za osobu koja koristi telekomunikacioni servis ili neki drugi sistem korišćen je naziv „korisnik”, bez želje da se implicira njena rodna pripadnost. Primetimo da se u novijoj literaturi na engleskom jeziku sve više govori o „korisnici”.

# Poglavlje 2

## Pronalaženje informacija

### 2.1 Uvod

Još pre 5000 godina, ljudi su shvatili da je za efikasno korišćenje informacija nužno da se one pravilno organizuju i arhiviraju. Već s pojavom prvih biblioteka, Sumerci su razvili sistem za klasifikovanje i pronalaženje tablica s tekstovima na klinastom pismu [lib07]. Od tada, na razvoj tehnika pronalaženja informacija ključan uticaj imala su dva pronalaska – štampa, čime je povećana produkcija informacija i (elektronski) računar, čime je omogućena obrada ove velike količine informacija. U literaturi se naglašava značaj vizionarskog rada V. Busha iz 1945. godine, u kome je predviđena upotreba računara za arhiviranje i pretraživanje celokupnog civilizacijskog znanja [Bus45].

Kao naučna disciplina, pronalaženje informacija dobija naročit zamah posle Drugog svetskog rata. Naime, hladni rat i trka u naoružanju doveli su do potrebe da se pomno prate naučne aktivnosti druge strane, naročito u oblastima nuklearne i raketne tehnike; stoga su SAD formirale posebne timove čiji je zadatak bio da prate sovjetske stručne publikacije i klasifikuju objavljene rade, kako bi se kasnije mogli lako pronaći. Danas, najznačajniji motivaciju razvoju tehnika pronalaženja informacija pružaju digitalne biblioteke i pretraživači interneta.

### 2.2 Definicija

Pronalaženje informacija (engl. *Information Retrieval*, IR) naučna je disciplina koja se bavi predstavljanjem, skladištenjem, organizacijom i pristupom informacionim objektima [Bae99]. U operativnom smislu, ono obuhvata:

- traženje informacija koje su sadržane u dokumentima,
- traženje samih dokumenata,

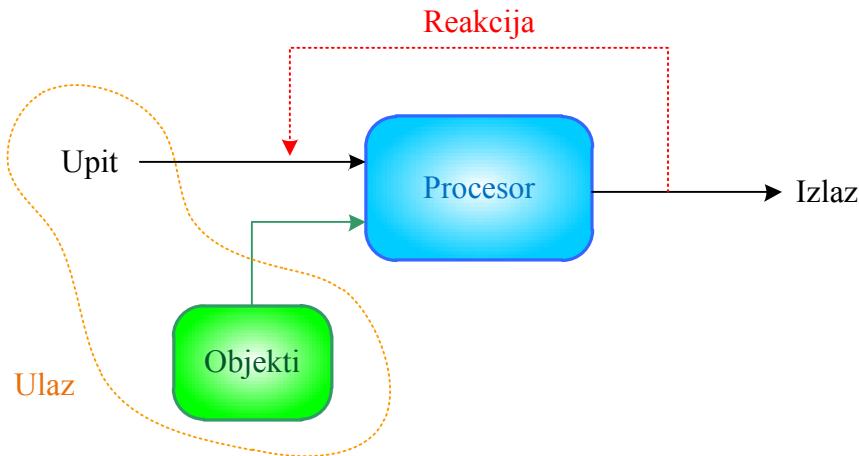
- traženje podataka o dokumentima (tzv. *metadata*),
- pretraživanje baza podataka itd.

U ovome kontekstu, izraz „informacija“ ima drugačije značenje od onog koje mu je dao Shannon, zasnivajući teoriju informacija. Ovde se informacija ne može kvantifikovati, već predstavlja podatak za sebe [Rij79]. Sistemi za pronalaženje informacija stoga svojim korisnicima ne pružaju nova znanja o materiji koja se odnosi na upit, već samo informišu korisnike o postojanju (ili nepostojanju) i lokaciji dokumenata koji su predmet upita.

Tehnike pronalaženja informacija prvo bitno su razvijane za potrebe klasifikovanja i pretraživanja tekstualnih dokumenata. Savremeno pronalaženje informacija je interdisciplinarna oblast, koja objedinjuje bibliotekarstvo, računarsku tehniku, matematiku (naročito statistiku), informacione tehnologije, lingvistiku, psihologiju, fiziku itd.

## 2.3 Osnovni elementi

Opšta blok-šema sistema za pronalaženje informacija data je na slici 2.1. Tri komponente sistema su ulaz, procesor i izlaz [Rij79].



Slika 2.1: Sistem za pronalaženje informacija.

Sistem ima dva ulaza u užem smislu i to korisničke upite i objekte. Upiti predstavljaju formalne izjave korisnika o informacijama koje se traže. U tom smislu, oni odražavaju korisnikovu trenutnu potrebu za informacijom. Objekti su surrogati ili reprezentacije originalnih dokumenata, koji se umesto njih čuvaju u bazi podataka sistema. Dobar primer za ovo nalazimo u bibliotekarstvu, gde se za kataloško predstavljanje bibliotečkih jedinica koriste indeksiranje preko ključnih reči i univerzalna decimalna klasifikacija (UDC). Primer nekih kategorija UDC za oblast hemije dati su u tabeli 2.1.

Upiti i objekti se predstavljaju u formi koja je pogodna za obradu u procesoru. Procesor je deo sistema koji obavlja pronalaženje. Proses pronalaženja obuhvata strukturiranje

Tabela 2.1: Izvod iz tabela univerzalne decimalne klasifikacije.

---

**5** Matematika. Prirodne nauke.

**54** Hemija. Kristalografija. Mineralogija.

**542** Praktična laboratorijska hemija. Preparativna i eksperimentalna hemija.

**542.1** Hemijske laboratorije.

**542.2** Laboratorijska oprema i instrumentacija uopšte. Laboratorijske metode uopšte.

**542.3** Merenje težine, mase. Merenje zapremine.

**542.4** Vruće i hladne aplikacije.

**542.5** Upotreba plamena. Lemilice.

**542.6** Rad s tečnostima.

**542.7** Rad s gasovima.

**542.8** Fizički, fizičko-hemijski i električni postupci.

**542.9** Hemijske reakcije. Specijalni hemijski procesi.

---

podataka sadržanih u upitima i objektima na pogodan način, npr. kroz klasifikovanje i samo pronalaženje u užem smislu.

Izlaz sistema za pronalaženje informacija tipično predstavlja lista naslova dokumenata koji se odnose na tekući upit.

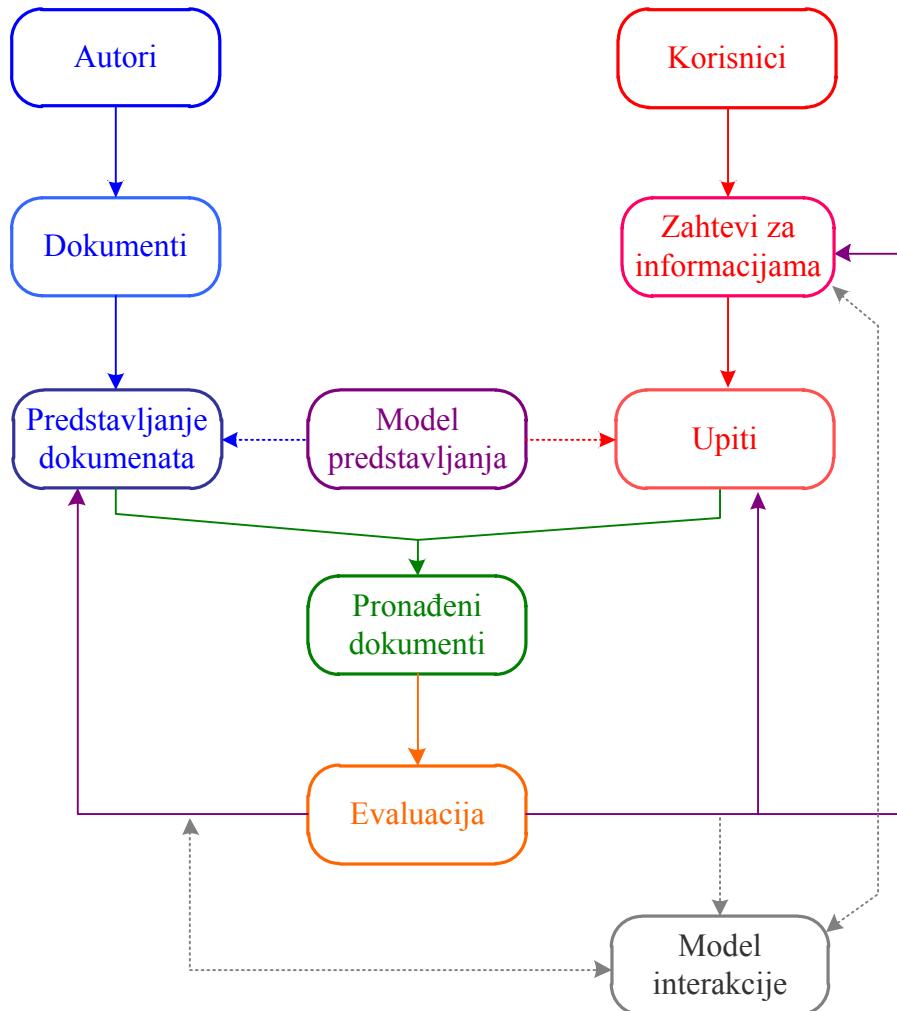
Tokom sesije, korisnik može menjati originalni upit, kako bi „pročistio“ dobijene rezultate. Ova procedura se naziva povratnom spregom ili reakcijom i na slici je simbolički prekazana isprekidanom crvenom linijom.

**Primer 2.1.** Korišćenjem [on-line tablica UDC](#), odredimo oblast na koju se odnosi šifra 547.

*Rešenje.* Zadata šifra odnosi se na organsku hemiju. ■

Proces pronalaženja informacija možemo detaljnije razraditi, čime dolazimo do šeme sa slike 2.2.

Autori prave dokumente koji se, prema usvojenom modelu, predstavljaju svojim surrogatima. Na drugoj strani su korisnici, koji imaju zahteve za informacijama. Ti zahtevi se, prema modelu predstavljanja i modelu interakcije, predstavljaju u vidu upita koji se postavljaju pred sistem za pronalaženje informacija. Na osnovu podataka sadržanih u surrogatima, sistem pronalazi dokumente koji korespondiraju korisničkim upitim. Korisnik u procesu evaluacije odvaja one pronađene dokumente koji se zaista odnose na njegovu potrebu za informacijom i, po potrebi, preduzima različite reakcije, npr. modifikuje predstavljanje dokumenata, upit, ili svoj prvobitni zahtev za informacijom.



Slika 2.2: Proces pronalaženja informacija – prema [Bae99], modifikovano.

## 2.4 Tehnike pronalaženja informacija

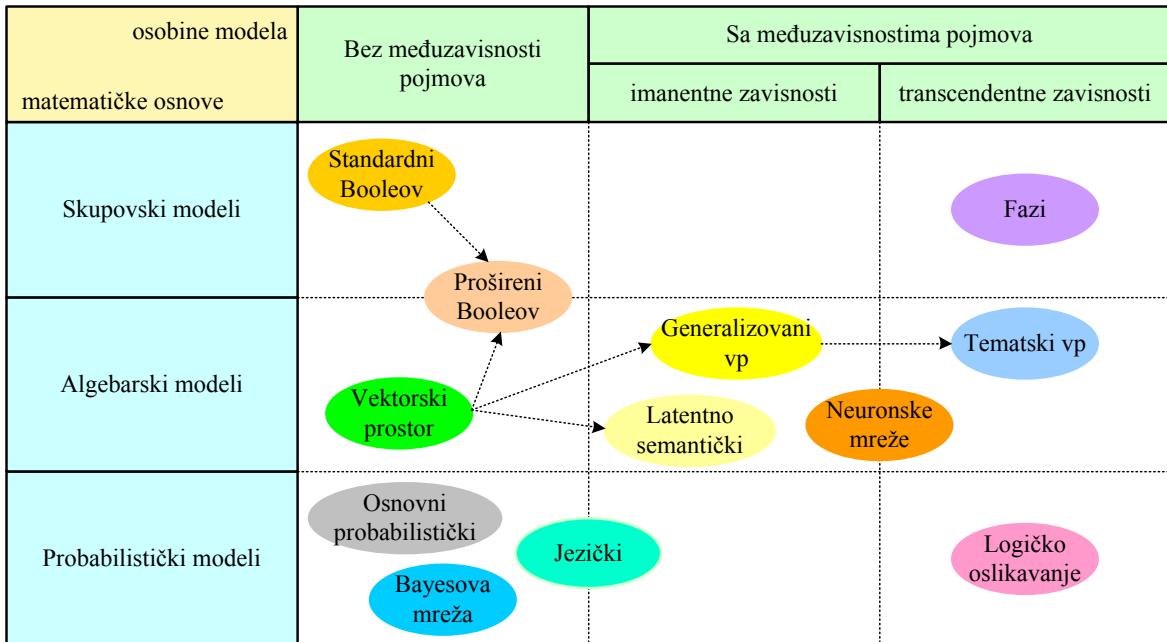
U literaturi je opisano više modela za pronalaženje informacija. U ovom odeljku, date su njihova klasifikacija i opšti pregled. Modeli koji se danas najviše koriste opisani su detaljnije u nastavku.

Formalno definisan, model pronalaženja informacija je uređena četvorka  $(D, Q, F, R)$ , u kojoj je:

- $D$  skup raspoloživih dokumenata, ili njihovih surogata (objekata),
- $Q$  skup upita koji odražavaju potrebe korisnika za informacijama,
- $F$  model predstavljanja dokumenata i upita i
- $R(q, d)$  funkcija rangiranja, koja svakom upitu  $q \in Q$  i dokumentu  $d \in D$  pridružuje realan broj.

Na slici 2.3, data je dvodimenzionalna klasifikacija modela za pronalaženje informacija, prema [Kur04]. Šemu ne treba shvatiti rigidno, jer, kao što se vidi, neki modeli

istovremeno pripadaju različitim kategorijama.



Slika 2.3: Klasifikacija modela za pronalaženje informacija prema [Kur04].

Vertikalnu dimenziju klasifikovanja čine matematičke osnove modela. Prema njoj, razlikujemo skupovske, algebarske i probabilističke modele.

U skupovskim modelima, dokumenti i upiti predstavljaju se skupovima svojih pojmljiva. Relevantnost dokumenta za tekući upit određuje se korišćenjem elemenata teorije skupova. Tipičan predstavnik ove grupe modela je tzv. standardni Booleov model, koji je bio prvi primjenjen za pronalaženje informacija. U ovome modelu i dokumenti i upiti se predstavljaju skupovima ključnih reči, koje su povezane logičkim operacijama *i*, *ili* i *ne*. Podudaranje između upita i dokumenta određuje se korišćenjem operacije jednakosti skupova, što znači da je posmatrani dokument relevantan za upit ako i samo ako se skup njegovih ključnih reči podudara s odgovarajućim skupom ključnih reči upita. Ovaj model se još primjenjuje u jednostavnijim pretraživačima. Ostali predstavnici su prošireni Booleov model [Sal83a] i fazi model [Cro94].

**Primer 2.2.** Data je tablica incidencije pojmljiva u Shakespeareovim dramama. Jedinačica na poziciji  $(t, d)$  označava da se pojmljivo  $t$  javlja u drami  $d$ , dok nula znači suprotno. Korišćenjem Booleovog modela, nađimo dokument koji odgovara upitu „ne Brut i Cezar”.

*Rešenje.* Dokumentima iz kolekcije pridružićemo binarne sekvene koje odgovaraju kolonama iz tablice incidencije:

*Julije Cezar:* 11110,  
*Bura:* 00000,  
*Hamlet:* 01100,

Tabela 2.2: Uz primer 2.2.

	Julije	Cezar	Bura	Hamlet	Othello	Macbeth
Antonije	1	0	0	0	1	
Brut	1	0	1	0	0	
Cezar	1	0	1	1	1	
Kalpurnija	1	0	0	0	0	
Kleopatra	0	0	0	0	0	

*Othello:* 00100 i

*Macbeth:* 10100.

Na isti način, upitu „ne Brut i Cezar” odgovara sekvenca  $x01xx$  („x” znači da je svejedno o kojoj se vrednosti radi). Izvršavanjem bit po bit ekvivalencije sekvene upita sa sekvencama dokumenata, zaključujemo da zadatom upitu odgovaraju dva dokumenta, *Othello* i *Macbeth*. ■

U algebarskim modelima, dokumenti i upiti se predstavljaju u vidu vektora, matrica ili, uopšteno, uređenih  $n$ -torki. Korišćenjem konačnog broja algebarskih operacija, oni se svode na skalarnu (jednodimenzionalnu) meru sličnosti. Najznačajniji predstavnik ove grupe modela je model vektorskog prostora [Sal75], iz koga su izvedeni generalizovani model vektorskog prostora [Won95], model tematskog vektorskog prostora [Bec03] i latentno semantički model [Dee90]. Ovde takođe spadaju i model zasnovan na neuronskim mrežama [Che95] i, delom, prošireni Booleov model.

U probabilističkim modelima, pronalaženje informacije se posmatra kao višestepeni statistički eksperiment. Primeri su osnovni probabilistički model [Mar60, Rob76], jezički model [Pon98], Bayesova mreža [Tur89] i logičko oslikavanje [Cre95].

Vertikalnu dimenziju klasifikovanja čine osobine modela, u smislu postojanja međuzavisnosti između pojmove kojima se opisuju dokumenti i upiti.

U nekim modelima se pretpostavlja da nema uzajamnih zavisnosti između pojmove. To se, na primer, u modelu vektorskog prostora izražava kroz ortogonalnost bazisnih vektora, a u probabilističkim modelima kroz pretpostavku nezavisnosti događaja.

U modelima u kojima postoje uzajamne zavisnosti između pojmove, one mogu biti imanentne i transcendentne. U prvom slučaju, stepen ove zavisnosti inherentno je određen parametrima modela, npr. brojem pojavljivanja određenih fraza u tekstu. U drugom slučaju, postoji spoljašnji izvor (npr. korisnik sistema) koji određuje stepen međuzavisnosti.

Od opisanih modela, najveću primenu u praksi našli su model vektorskog prostora, osnovni probabilistički model i model zasnovan na Bayesovim mrežama (mrežama za zaključivanje). Model na bazi fazi logike intenzivno se razvija i mogao bi u budućnosti steći veliki značaj. Pomalo iznenađuje činjenica da model na bazi neuronskih mreža ne

daje dobre rezultate za pronalaženje informacija [Bae99, Sin01]. U narednim odeljcima, stoga su detaljnije opisani probabilistički modeli i fazi model, dok je model vektorskog prostora, zbog svog izuzetnog značaja za personalizaciju, opisan u narednom poglavlju.

## 2.5 Osnovni probabilistički model

U osnovnom probabilističkom modelu, procesor treba da za svaki raspoloživi dokument i svaki postavljeni upit odgovori na sledeće pitanje:

Kolika je verovatnoća da je dokument relevantan za upit?

Prepostavlja se da je relevantnost dokumenta prema upitu (ili, bolje rečeno, zahtevu korisnika) binarna promenljiva, što znači da je dokument ili relevantan, ili nije. Takođe, prepostavlja se da relevantnost jednog dokumenta ne zavisi od karakteristika ostalih dokumenata.

Pitanje s početka ovoga odeljka postavlja se jer je cilj da se raspoloživi dokumenti rangiraju prema verovatnoćama svoje relevantnosti. U teoriji pronalaženja informacija, navodi se tzv. princip rangiranja verovatnoće, po kome je efikasnost sistema najveća onda kada se rezultati pretraživanja korisniku prezentuju uređeni po opadajućem redosledu verovatnoća relevantnosti. Korisnik tada može definisati pravilo odlučivanja po kome će prihvati ili odbaciti ponuđeni dokument.

Pošto stvarne verovatnoće relevantnosti nisu poznate, probabilistički model ih procenjuje iz nekih svojstava upita i dokumenta.

Definišimo uzajamno isključive događaje:

- $L$ : Dokument je relevantan,
- $\bar{L}$ : Dokument je irelevantan.

Neka je  $D$  posmatrani dokument i  $Q$  tekući upit<sup>1</sup>. Interesuje nas verovatnoća da je dokument  $D$  relevantan za upit  $Q$ ,  $P(L | D)$ . Pozivajući se na Bayesovu teoremu, možemo pisati

$$P(L | D) = \frac{P(D | L)P(L)}{P(D)}. \quad (2.1)$$

Na sličan način, za događaj  $\bar{L}$  je

$$P(\bar{L} | D) = \frac{P(D | \bar{L})P(\bar{L})}{P(D)}. \quad (2.2)$$

Formirajmo odnos ovih verovatnoća:

$$\frac{P(L | D)}{P(\bar{L} | D)} = \frac{P(D | L)P(L)}{P(D | \bar{L})P(\bar{L})}. \quad (2.3)$$

---

<sup>1</sup>Formalno bi trebalo pisati  $D_i$  i  $Q_j$ , ali ćemo izostaviti indekse radi jednostavnijeg zapisa.

Logaritmujmo sada obe strane ovoga izraza. Pošto je logaritamska funkcija uniformno rastuća, ovim se neće narušiti poredak neophodan za princip rangiranja. Tako dobijamo

$$\log \frac{P(L | D)}{P(\bar{L} | D)} = \log \frac{P(D | L)}{P(D | \bar{L})} + \log \frac{P(L)}{P(\bar{L})}. \quad (2.4)$$

Definišimo sada funkciju podudaranja,

$$M_1(D) = \log \frac{P(D | L)}{P(D | \bar{L})}. \quad (2.5)$$

Prema (2.4) je

$$M_1(D) = \log \frac{P(L | D)}{P(\bar{L} | D)} - \log \frac{P(L)}{P(\bar{L})}. \quad (2.6)$$

Drugi član na desnoj strani ove jednačine jednak je za sve dokumente. Ako bismo za svaki dokument mogli proceniti vrednost funkcije  $M_1(D)$ , stoga bismo mogli i uređiti dokumente po traženom poretku.

Dokumenti i upiti su opisani atributima, npr. ključnim rečima. Uvedimo sada važnu pretpostavku po kojoj su atributi uzajamno statistički nezavisni. Ova pretpostavka se naziva pretpostavkom nezavisnosti. Probabilistički model u kome važi pretpostavka nezavisnosti naziva se i naivnim Bayesovim modelom. Razmatranje opštijeg slučaja, u kome ne važi pretpostavka nezavisnosti, izašlo bi iz okvira ovoga kursa.

Ako s  $A_i$  označimo  $i$ -ti atribut, a s  $a_j$  njegovu vrednost za posmatrani dokument, tada možemo pisati

$$P(D | L) = \prod_{i,j} P(A_i = a_j | L), \quad (2.7)$$

$$P(D | \bar{L}) = \prod_{i,j} P(A_i = a_j | \bar{L}). \quad (2.8)$$

Funkcija podudaranja sada je data izrazom

$$M_1(D) = \sum_{i,j} \log \frac{P(A_i = a_j | L)}{P(A_i = a_j | \bar{L})}. \quad (2.9)$$

Odavde sledi da podudaranje dokumenta kao celine s upitom možemo izračunati ako odredimo podudaranja pojedinih njegovih atributa s tim upitom. U praktičnim primenama bilo bi pogodno da se iz razmatranja eliminišu oni atributi kojih nema u dokumentu, tj. čija je vrednost jednaka nuli. Definisaćemo, stoga, novu funkciju podudaranja,

$$M_2(D) = M_1(D) - \sum_i \log \frac{P(A_i = 0 | L)}{P(A_i = 0 | \bar{L})}. \quad (2.10)$$

Sada je

$$M_2(D) = \sum_{i,j} \left( \sum_{i,j} \log \frac{P(A_i = a_j | L)}{P(A_i = a_j | \bar{L})} - \log \frac{P(A_i = 0 | L)}{P(A_i = 0 | \bar{L})} \right) =$$

$$= \sum_{i,j} \log \frac{P(A_i = a_j | L)P(A_i = 0 | \bar{L})}{P(A_i = a_j | \bar{L})P(A_i = 0 | L)}. \quad (2.11)$$

Ovaj izraz možemo napisati u obliku

$$M_2(D) = \sum_{i,j} w(A_i = a_j), \quad (2.12)$$

gde je s  $w(A_i = a_j)$  označena težina koju atribut  $A_i$  ima kada mu je vrednost  $a_j$ ,

$$w(A_i = a_j) = \log \frac{P(A_i = a_j | L)P(A_i = 0 | \bar{L})}{P(A_i = a_j | \bar{L})P(A_i = 0 | L)}. \quad (2.13)$$

Primetimo da je  $w(A_i = 0) = 0$ , pa nulte vrednosti atributa ne utiču na vrednost funkcije podudaranja.

Izvedeni model može se značajno pojednostaviti ako pretpostavimo binarni skup vrednosti atributa, tj. da atribut  $A_i$  označava prisustvo ili odsustvo pojma  $t_i$ . Ako s  $p_i$  označimo uslovnu verovatnoću  $P(t_i | L)$ , a s  $\bar{p}_i$  verovatnoću  $P(t_i | \bar{L})$ , nakon nekoliko jednostavnih transformacija (2.13) postaje

$$w(A_i = a_j) = \log \frac{p_i(1 - \bar{p}_i)}{\bar{p}_i(1 - p_i)}. \quad (2.14)$$

Neka je ukupan broj raspoloživih dokumenata  $N$ . Neka je, dalje broj dokumenata koji su relevantni za tekući upit  $R$  i broj relevantnih dokumenata koji sadrže posmatrani pojam  $r$ . Odnosi između ovih parametara prikazani su u tabeli 2.3.

Tabela 2.3: Brojevi dokumenata u odnosu na tekući upit i posmatrani pojam.

	Relevantni za $Q$	Irelevantni za $Q$	Ukupno
Sadrže pojam $t_i$	$r_i$	$n_i - r_i$	$n_i$
Ne sadrže pojam $t_i$	$R - r_i$	$N - n_i - R + r_i$	$N - n_i$
Ukupno	$R$	$N - R$	$N$

Kada bi nam vrednosti ovih parametara bile poznate, važilo bi

$$p_i = \frac{r_i}{R} \quad (2.15)$$

i

$$\bar{p}_i = \frac{n_i - r_i}{N - R}. \quad (2.16)$$

Težina pojma tada bi bila data izrazom

$$w_i = \log \frac{r_i(N - n_i - R + r_i)}{(R - r_i)(n_i - r_i)}. \quad (2.17)$$

Aposteriorno znanje nije nam poznato u toku izvršavanja pretrage; zbog toga su predložene metode za procenu težine pojma koje koriste apriorno znanje o dokumentima, odnosno njihov ukupan broj,  $N$  i broj dokumenata koji sadrže posmatrani pojam,  $n_i$ .

U prvoj aproksimaciji, može se pretpostaviti da je broj relevantnih dokumenata,  $R$ , mnogo manji od ukupnog broja dokumenata u kolekciji,  $N$ . Stoga je

$$\bar{p}_i \approx \frac{n_i}{N}. \quad (2.18)$$

Dalje, pretpostavlja se da je verovatnoća  $p_i$  konstantna za sve pojmove. Tada je član  $\frac{p_i}{1-p_i}$  u (2.14) takođe konstantan i ne utiče na poredak dokumenata, pa se može izostaviti iz razmatranja. Pod ovim pretpostavkama, težina pojma konačno postaje

$$w_i = \log \frac{N - n_i}{n_i}, \quad (2.19)$$

pa je funkcija podudaranja

$$M_2(D) = \sum_i \log \frac{N - n_i}{n_i}. \quad (2.20)$$

Kao što ćemo kasnije videti, opisana aproksimacija umnogome podseća na takozvanu TF-IDF metodu, koja je primenjena u modelu vektorskog prostora.

U praktičnim primenama, često se koriste sledeće funkcije podudaranja:

$$PN(D) = \sum_{t \in Q, D} \frac{1 + \ln(tf)}{1 - s + s \frac{dl}{avdl}} qtf \ln \frac{N + 1}{df} \quad (2.21)$$

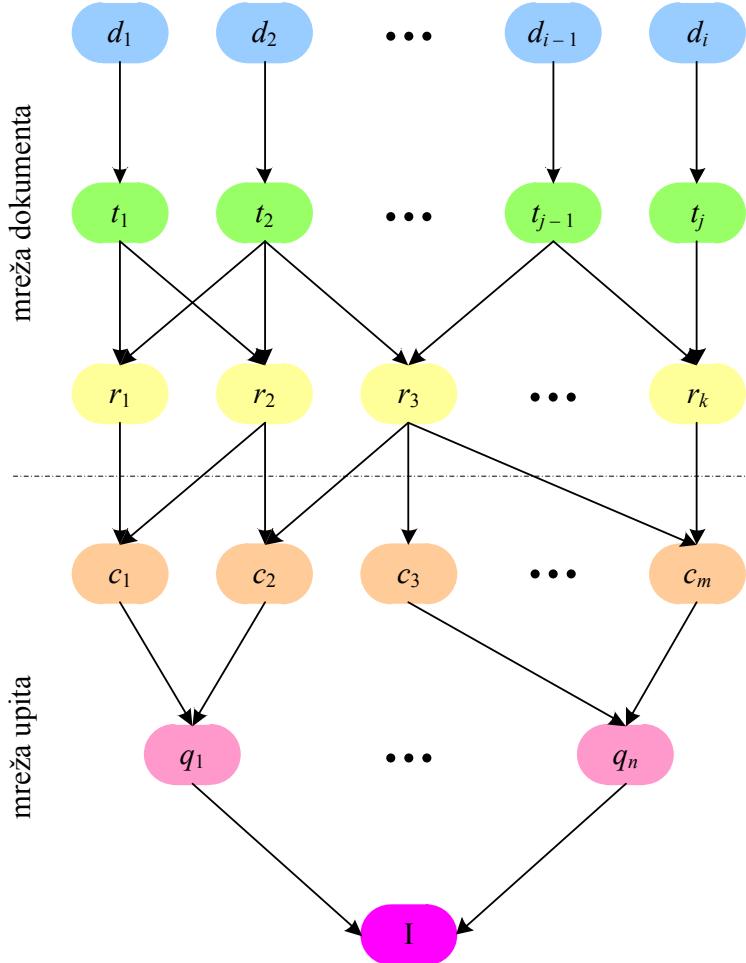
i

$$O(D) = \sum_{t \in Q, D} \ln \left( \frac{N - df + 0,5}{df + 0,5} \right) \frac{(k_1 + 1) tf}{k_1(1 - b) + b \frac{dl}{avdl} + tf} \frac{(k_3 + 1) qtf}{k_3 + qtf}. \quad (2.22)$$

U gornjim izrazima,  $tf$  je broj pojavljivanja posmatranog pojma  $t$  u tekućem dokumentu  $D$  (engl. *term frequency*);  $qtf$  je broj pojavljivanja pojma  $t$  u upitu  $Q$  (engl. *query term frequency*);  $N$  je, kao i ranije, ukupan broj dokumenata u kolekciji;  $df$  je broj dokumenata u kojima se javlja pojma  $t$  (engl. *document frequency*) – u prethodnim izrazima koristili smo oznaku  $n_i$ ;  $dl$  je broj reči u dokumentu  $D$  (engl. *document length*), dok je  $avdl$  prosečan broj reči u dokumentima iz kolekcije (engl. *average document length*). Sa  $s$ ,  $b$ ,  $k_1$  i  $k_3$  označene su konstante koje imaju ulogu parametara modela. Najčešće se usvajaju vrednosti  $s = 0,2$ ,  $b = 0,75$ ,  $k_1 \in [1, 2]$  i  $k_3 \in [0, 1000]$ .

## 2.6 Model Bayesovih mreža

Model Bayesovih mreža polazi od stanovišta da je pronalaženje informacija proces zaključivanja u kome se procenjuje verovatnoća da posmatrani dokument zadovoljava korisnikov zahtev za informacijama, koji je iskazan kroz tekući upit. Bayesova mreža za pronalaženje informacija, čija je najopštija struktura prikazana na slici 2.4, sastoji se iz mreže dokumenta i mreže upita. Obe mreže se sastoje iz čvorova, kojima je pridružena vrednost iz skupa {tačno, pogrešno} i grana.



Slika 2.4: Bayesova mreža za pronalaženje informacija.

Mreža dokumenta služi za predstavljanje raspoloživih dokumenata. Ova mreža je svojstvena datoj kolekciji dokumenata i njena struktura se ne menja tokom obrade upita. Sastoji se iz čvorova dokumenata,  $d$ , čvorova teksta,  $t$  i čvorova koncepta predstavljanja,  $r$ . Ako s  $D$  označimo skup dokumenata,  $T$  skup tekstova i  $R$  skup koncepcata predstavljanja i ako su njihovi kardinalni brojevi redom  $i, j$  i  $k$ , tada mreža dokumenta predstavlja prostor događaja  $S_d = D \times T \times R$ , gde je simbolom  $\times$  označen proizvod skupova. Pošto svaki čvor mreže može imati dve vrednosti, ukupan broj elemenata prostora  $S_d$  je  $2^{i+j+k}$ .

Svakom čvoru dokumenta odgovara jedan od raspoloživih dokumenata. Čvor dokumenta može se interpretirati kao događaj da se posmatra taj konkretni dokument. Čvoru dokumenta pridružena je apriorna verovatnoća posmatranja korespondirajućeg dokumenta. Opravdano je prepostaviti da je ova verovatnoća uniformno raspodeljena i da za svaki dokument iznosi  $1/i$ .

Čvor teksta odgovara određenom tekstu unutar dokumenta i, prema tome, može se interpretirati kao događaj da se posmatra neki konkretni tekst. Na ovom mestu treba dati važnu terminološku napomenu. Izraz „tekst“ ovde označava sadržaj dokumenta, koji, pored teksta u užem smislu, može obuhvatati i slike, audio, video itd. Izraz se

zadržao isključivo iz istorijskih razloga, pa ćemo ga i ovde koristiti.

Zavisnost teksta od dokumenta predstavlja se granom koja polazi od čvora dokumenta i završava se u čvoru teksta. Pretpostavićemo da postoji biunivoka korespondencija između čvorova dokumenta i čvorova teksta, tj. da svakom čvoru dokumenta odgovara tačno jedan čvor teksta i obrnuto. Kardinalni brojevi skupova  $D$  i  $T$  stoga su jednaki,  $i = j$ . Čvor teksta ima vrednost „tačno” samo onda kada se posmatra korespondirajući dokument.

Čvorovi koncepta predstavljanja odgovaraju atributima kojima se opisuju raspoloživi dokumenti, npr. ključnim rečima. Prisustvo atributa u dokumentu označava se strelicom od čvora teksta do čvora koncepta predstavljanja. Svakom čvoru koncepta predstavljanja odgovara uslovna verovatnoća posmatranja, koja zavisi od čvorova teksta koji su povezani s njim.

U grafovskoj terminologiji, mreža dokumenta predstavlja šumu, orijentisani aciklični graf s tri nivoa. Čvorovi dokumenta su korenovi, čvorovi teksta su unutrašnji čvorovi, a čvorovi koncepta predstavljanja su listovi šume. Čvorovi dokumenta imaju tačno po jedan čvor teksta kao dete, dok čvorovi teksta imaju po jedan ili više čvorova koncepta predstavljanja kao decu.

S druge strane, mreža upita predstavlja „invertovani” orijentisani aciklični graf, s jednim listom,  $I$ , koji odgovara tekućem zahtevu korisnika za informacijom, više unutrašnjih čvorova,  $q$ , koji predstavljaju upite kojima se formulišu zahtevi za informacijom i korenim čvorovima,  $c$ , koji odgovaraju konceptima kojima se iskazuju upiti. Mreža upita se konstruiše iznova za svaki zahtev za informacijom i njena se struktura menja tokom obrađivanja tog zahteva, npr. zbog reakcije koju korisnik preduzima da bi „pročistio” dobijene rezultate.

Označimo s  $C$  skup koncepata upita i s  $Q$  skup samih upita. Neka su njihovi kardinalni brojevi  $m$  i  $n$ , respektivno. Mreža upita tada predstavlja prostor događaja  $S_q = C \times Q \times I$ . Pokazuje se, ali ćemo taj dokaz ovde izostaviti, da se po cenu povećanja kompleksnosti mreže iz nje mogu eliminisati čvorovi upita, pa je  $|S_q| \leq 2^{m+1}$ . Ukupan prostor događaja koji se predstavlja Bayesovom mrežom za pronalaženje informacija je  $S_d \times S_q$ .

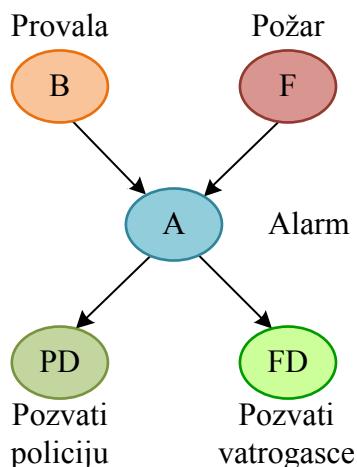
Koreni mreže upita su koncepti upita, primitive koje se koriste za iskazivanje zahteva za informacijom. Jedan čvor koncepta upita može imati više roditeljskih čvorova koncepta predstavljanja, što se predstavlja strelicama između njih. Čvoru koncepta upita pridružuje se verovatnoća kojom se opisuje uslovlenost posmatranog koncepta upita od roditeljskih koncepata predstavljanja. Čvorovi koncepata upita tako predstavljaju „most” između koncepata kojima se predstavljaju dokumenti i koncepata od kojih se grade upiti. U najjednostavnijem slučaju, koncepti upita podudaraju se s konceptima predstavljanja.

Čvor upita odgovara događaju da je zadovoljena posmatrana formulacija upita. Korisnikova potreba za informacijom u opštem slučaju nije precizno formulisana. Tu,

implicitnu potrebu, pokušavamo učiniti eksplisitnom tako što ju izražavamo preko jednog ili više upita koji imaju formalnu interpretaciju. Malo je verovatno da će neki od ovih upita sam za sebe u potpunosti odgovarati tekućem zahtevu za informacijom; da bi se taj zahtev što bolje formulisao, u mreži se stoga može nalaziti više čvorova upita.

Čvor I odgovara događaju da je zadovoljen korisnikov zahtev za informacijom. U opštem slučaju, ne može se tvrditi da će se ovaj događaj realizovati za proizvoljan skup dokumenata. Zadatak Bayesove mreže je da, kroz kombinovanje informacije o sadržaju raspoloživih dokumenata i informacije o upitima, proceni verovatnoću njegove realizacije. Ako su poznate apriorne verovatnoće izbora pojedinih dokumenata koji su koreni mreže i uslovne verovatnoće povezane s unutrašnjim čvorovima, moguće je odrediti aposteriornu verovatnoću da posmatrani dokument zadovoljava tekući zahtev za informacijom.

**Primer 2.3.** Na primeru sledeće Bayesove mreže, pokažimo kako se računa aposteriorna verovatnoća, npr. došlo je do provale ukoliko su pozvani policija i vatrogasci.



Slika 2.5: Uz primer 2.3.

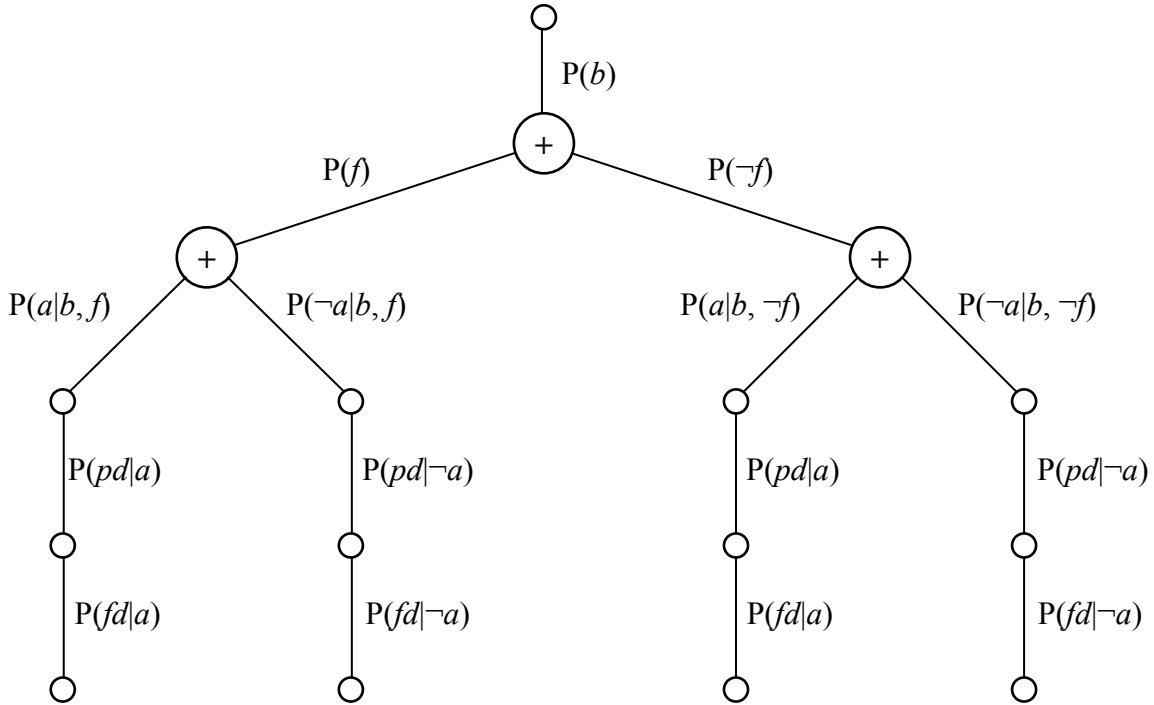
*Rešenje.* Po definiciji uslovne verovatnoće je

$$P(B | PD, FD) = \frac{P(B, PD, FD)}{P(PD, FD)}.$$

Zadržimo se na verovatnoći iz brojnika,  $P(B, PD, FD)$ . Po formuli totalne verovatnoće, ona je data s

$$\begin{aligned} P(B, PD, FD) &= \sum_f \sum_a P(B, f, a, PD, FD) = \\ &= P(B) \sum_f P(f) \sum_a P(a | B, f) P(PD | a) P(FD | a). \end{aligned}$$

Grafički prikaz izračunavanja ove verovatnoće ilustrovan je na narednoj slici. S nje vidimo ne samo to da je postupak prilično zametan, već i da se neki koraci izračunavanja ponavljaju.



Slika 2.6: Izračunavanje verovatnoće u primeru 2.3.

Po cenu nešto duže pripreme, do rezultata se brže može doći tzv. postupkom eliminacije promenljivih:

$$\begin{aligned}
 P(B, PD, FD) &= P(B) \sum_f P(f) \sum_a P(a | B, f) P(PD | a) P(FD | a) = \\
 &= P(B) \sum_f P(f) \sum_a P(a | B, f) P(PD | a) f_{FD}(a) = \\
 &= P(B) \sum_f P(f) \sum_a P(a | B, f) f_{PD}(a) f_{FD}(a) = \\
 &= P(B) \sum_f P(f) \sum_a f_A(a, b, f) f_{PD}(a) f_{FD}(a) = \\
 &= P(B) \sum_f P(f) f_{\bar{A}, PD, FD}(b, f) = \\
 &= P(B) f_{\bar{F}, \bar{A}, PD, FD}(b) = \\
 &= f_B(b) f_{\bar{F}, \bar{A}, PD, FD}(b).
 \end{aligned}$$

Ideja je, dakle, da se relevantni međurezultati pamte i kasnije koriste u izračunavanju. ■

## 2.7 Fazi model

Osnovna razlika fazi skupova u odnosu na „obične” je u tome što „običnim” skupovima neki element ili pripada, ili ne pripada; kod fazi skupova, govori se o stepenu pripadnosti elementa skupu. To znači da posmatrani element može pripadati jednom skupu u potpunosti, delimično, ili mu ne pripadati uopšte.

Neka je dat prostor događaja  $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ . Fazi skup  $A$ , kao njegov podskup, definiše se funkcijom pripadanja  $\mu_A : \Omega \mapsto [0, 1]$ , pri čemu  $\mu_A(\omega_i)$  predstavlja stepen pripadnosti elementa  $\omega_i$  skupu  $A$ :

$$A = \left\{ \frac{\mu_A(\omega)}{\omega} \mid \omega \in \Omega \right\}. \quad (2.23)$$

Osnovne operacije nad fazi skupovima date su u tabeli 2.4.

Tabela 2.4: Operacije nad fazi skupovima.

Oznaka	Značenje	Definicija
$A \subseteq B$	$A$ je podskup $B$	$\forall \omega \in \Omega, \mu_A(\omega) \leq \mu_B(\omega)$
$A = B$	jednakost skupova	$\forall \omega \in \Omega, \mu_A(\omega) = \mu_B(\omega)$
$\bar{A}$	komplement skupa	$\mu_{\bar{A}} = 1 - \mu_A(\omega)$
$A \cap B$	presek skupova	$\mu_{A \cap B} = \min(\mu_A(\omega), \mu_B(\omega))$
$A \cup B$	unija skupova	$\mu_{A \cup B} = \max(\mu_A(\omega), \mu_B(\omega))$

**Primer 2.4.** Dati su fazi skupovi  $A = \left\{ \frac{0,8}{1}, \frac{0,35}{p}, \frac{0,7}{X} \right\}$  i  $B = \left\{ \frac{0,3}{1}, \frac{0,5}{p}, \frac{0,1}{\kappa} \right\}$ . Odredimo vrednost izraza  $A \cup \bar{B}$ .

*Rešenje.* Primetimo da je  $\Omega = \{1, p, X, \kappa\}$ . Prema tabeli 2.4 je  $\bar{B} = \left\{ \frac{0,7}{1}, \frac{0,5}{p}, \frac{1}{X}, \frac{0,9}{\kappa} \right\}$ . Stoga je  $A \cup \bar{B} = \left\{ \frac{0,8}{1}, \frac{0,5}{p}, \frac{1}{X}, \frac{0,9}{\kappa} \right\}$ . ■

Primenjen na pronalaženje informacija, fazi model prepostavlja da je svaki dokument donekle relevantan za tekući upit, pri čemu treba odrediti stepen relevantnosti.

Fazi model za pronalaženje informacija definiše se kao uređena četvorka  $(T, D, Q, \mu_F)$ .

$S T = \{t_1, t_2, \dots, t_n\}$  označen je skup ključnih reči kojima se opisuju dokumenti i upiti.

Skup raspoloživih dokumenata je  $D = \{D_1, D_2, \dots, D_k\}$ . Dokument  $D_i \in D, i = 1, 2, \dots, k$ , predstavlja fazi skup ključnih reči,

$$D_i = \left\{ \frac{\mu_D(t)}{t} \mid t \in T \right\}, \quad (2.24)$$

pri čemu se često koristi sledeća funkcija pripadanja:

$$\mu_D(t) = \frac{n_t}{\max_{t_j}(n_{t_j})arch_t}. \quad (2.25)$$

Ovde je s  $n_t$  označen broj pojavljivanja ključne reči  $t$  u dokumentu  $D_i$ , a s  $arch_t$  ukupan broj pojavljivanja te ključne reči u svim raspoloživim dokumentima.

Skup upita označen je s  $Q = \{Q_1, Q_2, \dots, Q_l\}$ . Svaki upit je takođe fazi skup ključnih reči,

$$Q_i = \left\{ \frac{\mu_Q(t)}{t} \mid t \in T \right\}. \quad (2.26)$$

Funkcija pripadanja sada označava značaj ključne reči  $t$  za tekući upit. Složeniji upiti grade se od jednostavnijih primenom operacija unije, preseka i komplementiranja skupova, prema tabeli 2.4.

Konačno,  $\mu_F$  je funkcija pripadanja,  $\mu_F : D \times Q \rightarrow [0, 1]$ , koja određuje stepen u kome dokument  $D_i$  zadovoljava tekući upit  $Q_j$ . Jedan od njenih mogućih analitičkih oblika je

$$\mu_F = \sum_{t \in T} \frac{\min(\mu_D(t), \mu_Q(t))}{\mu_Q(t)}. \quad (2.27)$$

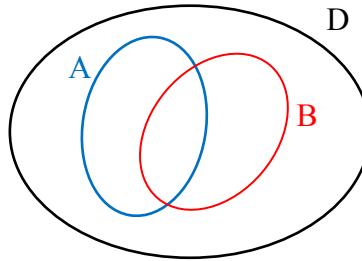
Sistem računa ovu vrednost za svaki raspoloživi dokument i, na kraju, može da korisniku kao relevantne ponudi samo one dokumente za koje je  $\mu_F$  veće od unapred zadatog praga.

## 2.8 Performanse sistema za pronalaženje informacija

Performanse sistema za pronalaženje informacija mogu se iskazati kroz više pokazatelja, kao što su [Rij79]:

- pokrivanje skupa raspoloživih dokumenata, tj. „dubina” do koje sistem ide u traženju relevantnih činjenica,
- kašnjenje, tj. prosečan interval vremena između postavljanja upita i dobijanja odgovora,
- vid prezentacije odgovora,
- napor koji korisnik mora da uloži da bi dobio odgovor,
- sposobnost pronalaženja, tj. deo relevantnog materijala koji sistem nalazi kao odgovor na upit,
- preciznost sistema, tj. deo pronađenog materijala koji je zaista relevantan za tekući upit itd.

Uobičajeno je da se kao mera efikasnosti sistema, u smislu zadovoljavanja potreba korisnika, koriste poslednja dva faktora, preciznost i sposobnost pronalaženja.



Slika 2.7: Skupovi dokumenata.

Neka je  $D$  skup svih raspoloživih dokumenata. Označimo sa  $A$  skup relevantnih dokumenata, a sa  $B$  skup pronađenih dokumenata, kao što je to prikazano na slici 2.7. Posmatrajmo tabelu 2.5.

Tabela 2.5: Skupovi dokumenata.

	Relevantni za $Q$	Irelevantni za $Q$	Ukupno
Pronađeni	$A \cap B$	$\bar{A} \cap B$	$B$
Nisu pronađeni	$A \cap \bar{B}$	$\bar{A} \cap \bar{B}$	$\bar{B}$
Ukupno	$A$	$\bar{A}$	$D$

Preciznost se definiše kao odnos broja relevantnih dokumenata među pronađenima i ukupnog broja pronađenih dokumenata:

$$PRE = \frac{|A \cap B|}{|B|}. \quad (2.28)$$

Sposobnost pronalaženja se iskazuje kroz odnos broja relevantnih dokumenata među pronađenima i ukupnog broja relevantnih dokumenata,

$$REC = \frac{|A \cap B|}{|A|}. \quad (2.29)$$

Iako nije poznat analitički oblik njihove međuzavisnosti, preciznost i sposobnost pronalaženja po svemu sudeći nisu nezavisne veličine. Na današnjem stepenu razvoja tehnike, veća preciznost se postiže po cenu manje sposobnosti pronalaženja informacija i obrnuto.

Istaknimo da je relevantnost u osnovi subjektivan pojam. Različiti korisnici mogu da različito doživljavaju relevantnost (ili irelevantnost) dokumenta za zadati upit; ova razlika, međutim, nije toliko značajna da bi dovela u pitanje rezultate eksperimenata kojima se ocenjuju performanse sistema za pronalaženje informacija [Rij79, Bae99]. Naime, uobičajena eksperimentalna praksa je da se skupovi raspoloživih dokumenata (tzv. kolekcije) daju na uvid korisnicima koji su stručnjaci za disciplinu na koju se ti dokumenti odnose i koji za svaki dokument određuju je li relevantan za postavljeni upit, ili nije. Ovako markirani dokumenti potom se prepuštaju sistemu za pronalaženje informacija; na taj način, pri eksperimentalnom testiranju, za svaki upit unapred znamo koji su dokumenti relevantni. Motivacija za ovakav pristup leži u očekivanju da će

sistem za pronalaženje informacija koji dobro radi pod eksperimentalnim uslovima takođe imati dobre performanse i pod operativnim uslovima u praksi, gde relevantnost pojedinih dokumenata nije apriorno poznata.

**Primer 2.5.** Dati su rezultati pretrage za upit „etf”:

*pronađeno:*

1. Exchange-traded fund (Wikipedia)
2. Exchange-traded fund (Investopedia)
3. European Training Foundation
4. Exchange-Traded Funds Center (Yahoo! Finance)
5. Univerzitet u Beogradu, Elektrotehnički fakultet

*nije pronađeno:*

6. Exchange Traded Funds Research (Morningstar)
7. Department of Employee Trust Funds Wisconsin
8. Family of Exchange Traded Funds (iShares)
9. European Transport Workers' Foundation
10. The Alternative ETF Company (ProShares)
11. Environmental Technologies Fund
12. ETF Group Nederweert Holland
13. ETF Acoustics Canada
14. Exhibitions and Trade Fairs
15. Elektrotehnički fakultet Podgorica

Odredimo preciznost i sposobnost pronalaženja za dve pretpostavljene interpretacije upita, „elektrotehnički fakultet” i „exchange-traded fund”.

*Rešenje.* Ukoliko je značenje upita „elektrotehnički fakultet”, relevantni su dokumenti s rangovima 5 i 15, od kojih je pronađen samo onaj pod brojem 5. Preciznost je, zato,  $1/5$ , a sposobnost pronalaženja  $1/2$ .

U drugom slučaju, u kolekciji postoji šest relevantnih dokumenata (1, 2, 4, 6, 8 i 10) od kojih su pronađena tri (1, 2 i 4). Preciznost je sada  $3/5$ , a sposobnost pronalaženja  $3/6 = 1/2$ , isto kao i u prvom slučaju. ■

U literaturi se navode i druge mere performansi, koje ćemo navesti radi kompletnosti prikaza.  $F_\alpha$ -mera data je izrazom

$$F_\alpha = (1 + \alpha) \frac{PRE \cdot REC}{\alpha \cdot PRE + REC}, \quad (2.30)$$

gde je  $\alpha$  nenegativan realni broj. U praksi se najčešće koriste mere  $F_{0,5}$ ,  $F_1$  i  $F_2$ .

Koeficijent ispadanja predstavlja odnos broja irelevantnih dokumenata među pronađenima i ukupnog broja irelevantnih dokumenata,

$$FALL = \frac{|\bar{A} \cap B|}{|\bar{A}|}. \quad (2.31)$$

S preciznošću i sposobnošću pronalaženja povezan je relacijom

$$PRE = \frac{REC \cdot G}{REC \cdot G + FALL(1 - G)}, \quad (2.32)$$

gde je  $G$  generalnost, udeo relevantnih dokumenata u čitavoj kolekciji,

$$G = \frac{|A|}{|D|}. \quad (2.33)$$

Posebnu grupu mera performansi sistema za pronalaženje informacija čine tzv. mere asocijacija.

Koeficijent poklapanja odgovara broju pronađenih relevantnih dokumenata

$$C_m = |A \cap B|. \quad (2.34)$$

Diceov koeficijent dat je izrazom

$$C_D = 2 \frac{|A \cap B|}{|A| + |B|}, \quad (2.35)$$

a Jaccardov koeficijent

$$C_J = \frac{|A \cap B|}{|A \cup B|}. \quad (2.36)$$

Kosinusni koeficijent dat je izrazom

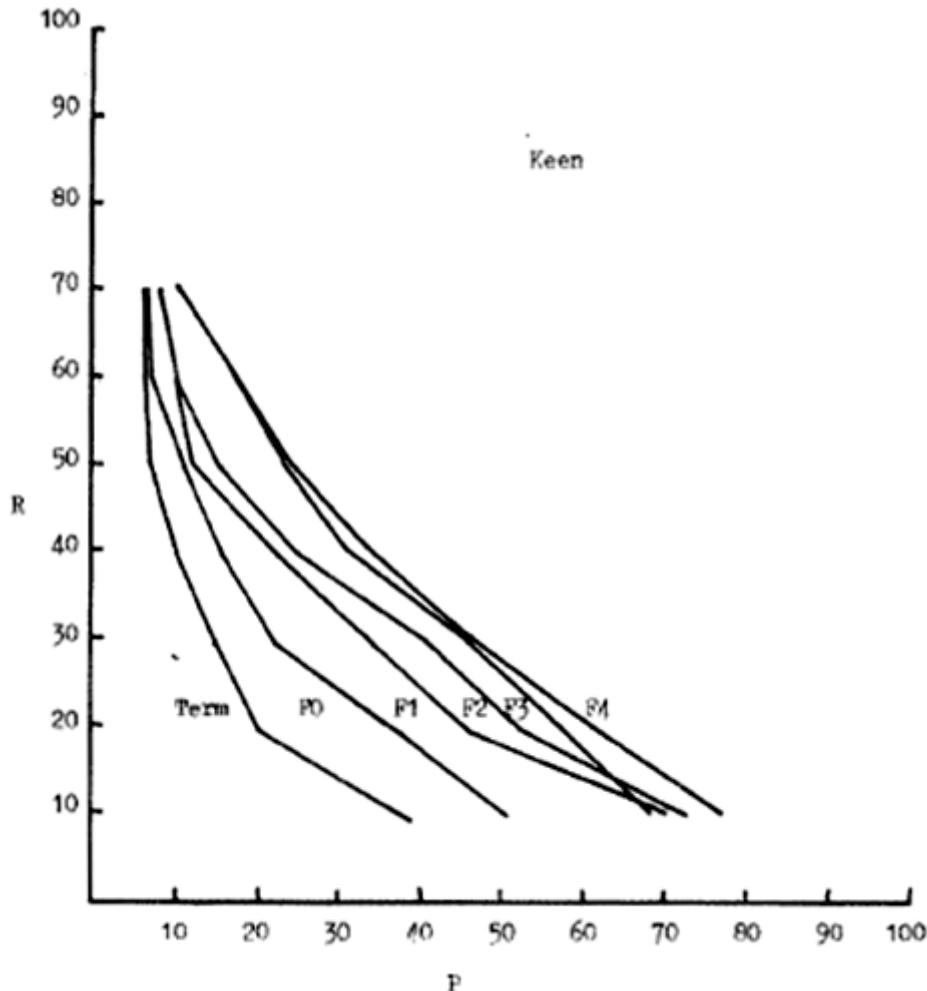
$$C_{\cos} = \frac{|A \cap B|}{|A|^{1/2} |B|^{1/2}}. \quad (2.37)$$

Koeficijent preklapanja je

$$C_{ol} = \frac{|A \cap B|}{\min(|A|, |B|)}. \quad (2.38)$$

Pitanje koje se sada logično postavlja je koji je metod pronalaženja informacija najbolji. Odgovor je jednostavan: Ne postoji metod koji bi imao najbolje performanse za sve kolekcije dokumenata i za sve upite. Čak i jedan model, primjenjen nad jednom kolekcijom dokumenata, postiže različite performanse za različite upite. Dobar primer za to pruža klasičan rezultat Robertsona i Spärck Jonesove, koji je prikazan na slici 2.8.

Iako postoji više „podjednako” dobrih modela za pronalaženje informacija, jedan je naročito pogodan za kombinovanje s tehnikama modeliranja korisnika u cilju personalizacije telekomunikacionih servisa. To je model vektorskog prostora, koji je detaljno opisan u narednoj glavi.



Slika 2.8: Performanse osnovnog probabilističkog modela, prema [Rob76]. Na apscisi je preciznost, a na ordinati sposobnost pronalaženja, obe u %.

## 2.9 Pitanja i zadaci

**2.1.** Objasnite princip rangiranja verovatnoće.

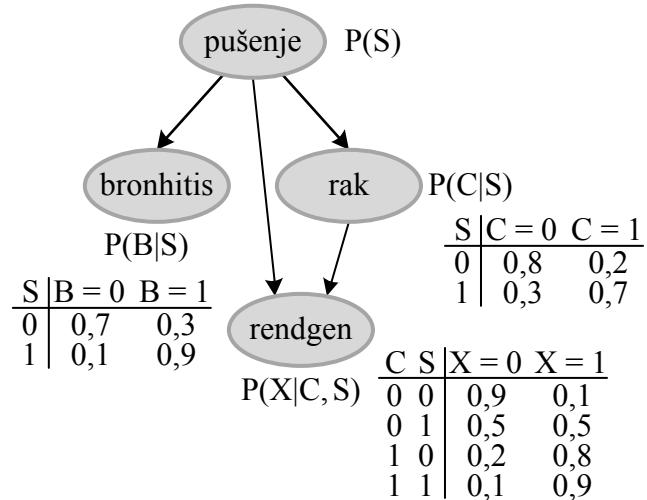
**2.2.** Za zadati rang  $k$ , neka je  $R_k$  broj relevantnih dokumenata čiji su rangovi od 1 do  $k$ . Ako sistem za pronalaženje informacija računa verovatnoću da je dokument  $D_i$  relevantan za upit  $Q$ ,  $P(R | D_i, Q)$ , tada je očekivana vrednost  $R_k$

$$ER_k = \sum_{i=1}^k P(R | D_i, Q).$$

- a) Svođenjem na absurd pokažite da se ovo očekivanje maksimizira ako su dokumenti rangirani po opadajućem redosledu verovatnoća.

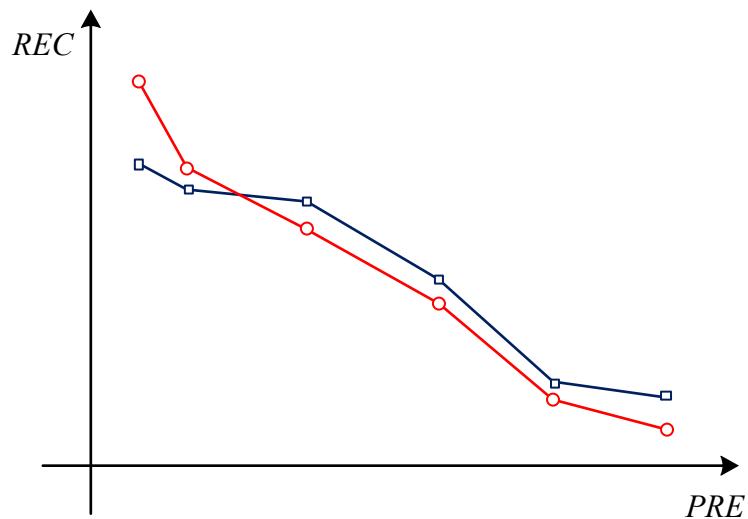
- b) Ako je broj relevantnih dokumenata u kolekciji  $\mathcal{R}$ , čemu su u ovome slučaju jednake očekivane vrednosti preciznosti i sposobnosti pronalaženja?
- c) Pokažite da se rezultat prethodne tačke maksimizira ako su dokumenti rangirani po opadajućem redosledu verovatnoća.

**2.3.** Na slici je data Bayesova mreža s naznačenim uslovnim verovatnoćama prelaska.



Kolika je verovatnoća da je pacijent koji je upućen na rendgenski pregled pušač?

**2.4.** Na slici su prikazane krive zavisnosti preciznosti od sposobnosti pronalaženja za dva sistema za pretraživanje stručnih tekstova.



Koji biste sistem izabrali ukoliko želite da pronađete sve tekstove koji se odnose na neku oblast? Obrazložite odgovor.

**2.5.** Sistem za pronalaženje informacija kao odgovor na upit vraća listu 10 rangiranih dokumenata. Neka je za taj upit stvarno relevantno 5 dokumenata, pri čemu su relevantni među pronađenima oni s rangovima 2, 3, 4 i 8. Konstruišite grafik zavisnosti preciznosti od sposobnosti pronalaženja za ovaj upit, za sledeći skup vrednosti sposobnosti pronalaženja:  $\{0,0, 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1,0\}$ .

**2.6.** Neka je upit u sistem za pronalaženje informacija „ckd” i neka su rezultati:

Pronađeno:	Nije pronađeno:
chronic kidney disease	count key data
complete knock down	ČKD Dopravní Systémy
cyclical ketogenic diet	acute kidney disease
ČKD Praha, a.s.	RT6N1

- a) Odredite preciznost i sposobnost pronalaženja za dva pretpostavljena značenja upita.  
 b) Koliko iznosi  $F$ -mera za slučajeve a), ako se značaj preciznosti vrednuje dvostruko više od značaja sposobnosti pronalaženja?

# Poglavlje 3

## Model vektorskog prostora

### 3.1 Uvod

U ovom poglavlju, objašnjen je model vektorskog prostora (engl. *Vector Space Model*, VSM), koji se široko koristi u oblasti pronalaženja informacija.

Mogućnost primene modela vektorskog prostora za pronalaženje informacija prvi je uvideo Salton u radu [Sal63], koji razmatra pretraživanje teksta. Zanimljivo je da mnogi autori zasnivanje VSM pogrešno referenciraju za jedan nepostojeći Saltonov rad iz 1975. godine, što je detaljno komentarisano u [Dub04]. Originalna ideja je razrađena u [Sal75], a formalnu definiciju VSM za pronalaženje informacija srećemo u [Sal83b] i [Sal89].

Suština modela vektorskog prostora je da se dokumenti (nekada tekstualni dokumenti, bibliotečke jedinice itd, a danas multimedijalni dokumenti i telekomunikacioni servisi) predstavljaju vektorom svojih odlika. Definicija odlike nije inherentna u modelu, već se kao odlike dokumenta tipično koriste ključne reči ili fraze. Na taj način, dokument se opisuje skupom ključnih reči, pa svaka reč u rečniku postaje nezavisna dimenzija visokodimenzionalnog vektorskog prostora. Ako posmatranom dokumentu pripada neka ključna reč, tada se u njegovom vektoru na poziciji te ključne reči nalazi nenulta vrednost. Detaljan matematički opis modela vektorskog prostora dat je u nastavku.

### 3.2 Opis

Počećemo formalnom definicijom vektorskog prostora. Neka su dati skup  $\Omega$  i polje  $(\mathbb{F}, +, \cdot)$ . Za skup  $\Omega$  kaže se da predstavlja *vektorski prostor* nad poljem  $\mathbb{F}$ , ako je zatvoren nad operacijom sabiranja,

$$(\forall \mathbf{x}, \mathbf{y} \in \Omega) \quad \mathbf{x} + \mathbf{y} \in \Omega \tag{3.1}$$

i množenja elementima polja  $\mathbb{F}$ ,

$$(\forall \alpha \in \mathbb{F}) (\forall \mathbf{x} \in \Omega) \quad \alpha \mathbf{x} \in \Omega \quad (3.2)$$

i ako za sve elemente  $\mathbf{x}, \mathbf{y}$  i  $\mathbf{z} \in \Omega$  i  $\alpha$  i  $\beta \in \mathbb{F}$  važe sledeće aksiome:

$$\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}, \quad (3.3)$$

$$\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}, \quad (3.4)$$

$$(\exists \mathbf{0} \in \Omega) \quad \mathbf{0} + \mathbf{x} = \mathbf{x}, \quad (3.5)$$

$$(\exists (-\mathbf{x}) \in \Omega) \quad \mathbf{x} + (-\mathbf{x}) = \mathbf{0}, \quad (3.6)$$

$$1\mathbf{x} = \mathbf{x}, \quad (3.7)$$

$$\alpha(\beta\mathbf{x}) = (\alpha\beta)\mathbf{x}, \quad (3.8)$$

$$\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}, \quad (3.9)$$

$$(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}. \quad (3.10)$$

Elementi vektorskog prostora nazivaju se vektorima, a elementi polja  $\mathbb{F}$  skalarima. Aksiome (3.3)-(3.6) pokazuju da je  $\Omega$  komutativna ili Abelova grupa u odnosu na operaciju sabiranja vektora. Ostale aksiome odnose se na množenje vektora skalarom.

Primetimo da postojanje vektorskog prostora implicira linearost sistema. Bilo koja dva elementa sistema mogu se sabrati i, takođe, bilo koji element sistema može se pomnožiti konstantom, da bi se dobio novi element sistema.

Osnovna postavka modela vektorskog prostora u teoriji pronalaženja informacija je da se dokumenti predstavljaju vektorima odlika. Ključne reči po kojima se dokumenti indeksiraju čine bazu vektorskog prostora. Vrednost  $i$ -te koordinate vektora odlike odgovara „sadržaju“  $i$ -te ključne reči u posmatranom dokumentu. To može biti binarna promenljiva, pri čemu vrednost 0 označava odsustvo, a vrednost 1 prisustvo ključne reči u dokumentu, ili, pak, kontinualna promenljiva, koja označava „količinu“ ključne reči u dokumentu.

Sistemi za pronalaženje informacija tipično koriste samo pozitivni hiperkvadrant vektorskog prostora, što znači da vektori odlika ne sadrže negativne vrednosti [Sin01], mada je konzistentnost modela očuvana i u tom slučaju [Rag86]. Pri svakom korisničkom upitu, sistem formira vektor upita i poredi ga s vektorima odlika raspoloživih dokumenata tako što računa njihovu uzajamnu sličnost, prema usvojenoj metodi. Na kraju, korisniku se nudi onaj dokument koji ima najveću sličnost s upitom.

Neka su  $t_i$ ,  $i = 1, \dots, n$  ključne reči kojima se opisuju dokumenti. Prepostavimo dalje da svakoj ključnoj reći  $t_i$  odgovara vektor  $\mathbf{t}_i$ . Bez umanjenja opštosti, usvojićemo da su ovi vektori jediničnog modula. Svaki raspoloživi dokument  $D_j$ ,  $j = 1, \dots, m$ , sada predstavljamo vektorom  $\mathbf{D}_j = (d_{1j}, d_{2j}, \dots, d_{nj})$ , čije su koordinate realni brojevi koji odražavaju zastupljenost ili značaj pojedinih ključnih reči u tome dokumentu. Na ovaj

način, ključne reči formiraju generišući skup vektorskog prostora dokumenata. Svaki vektor u ovome prostoru može se predstaviti linearom kombinacijom ključnih reči, tj.

$$\mathbf{D}_j = \sum_{i=1}^n d_{ij} \mathbf{t}_i, \quad j = 1, \dots, m. \quad (3.11)$$

Minimalan generišući skup čini bazu vektorskog prostora.

Za vektore  $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_k$  kažemo da su linearno zavisni ako postoje skalari  $a_1, a_2, \dots, a_k$ , koji nisu svi jednaki nuli, takvi da važi

$$\sum_{i=1}^k a_i \mathbf{y}_i = \mathbf{0}. \quad (3.12)$$

U suprotnom, vektori su linearno nezavisni.

Sada ćemo navesti neke važne definicije i osobine vektorskih prostora, koje su poznate iz linearne algebre [Nic93]:

- Ako generišući skup vektorskog prostora  $\Omega$  čini  $n$  vektora  $\mathbf{t}_i$ ,  $i = 1, \dots, n$ , tada svaki skup linearne nezavisnosti vektora u prostoru  $\Omega$  sadrži najviše  $n$  vektora.
- Bazu vektorskog prostora čini generišući skup čiji su elementi linearne nezavisnosti vektori. Stoga baza prostora  $\Omega$  ima najviše  $n$  vektora, pa njegova dimenzija nije veća od  $n$ .
- Iz generišućeg skupa vektorskog prostora uvek je moguće konstruisati bazu tako što se iz njega eliminišu vektori koji su linearne kombinacije drugih.
- Za zadatu bazu  $\{\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_l\}$ ,  $l \leq n$ , svaki vektor  $\mathbf{x} \in \Omega$  može se na jedinstven način predstaviti kao linearna kombinacija elemenata baze,

$$\mathbf{x} = \sum_{i=1}^l c_i \mathbf{t}_i, \quad (3.13)$$

pri čemu je  $l$  dimenzija prostora  $\Omega$ .

- Ako je jedna baza prostora  $\Omega$   $\{\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_l\}$ ,  $l \leq n$ , tada će svaki skup  $l$  linearne nezavisnosti vektora iz  $\Omega$  takođe činiti bazu vektorskog prostora, čija je dimenzija ponovo  $l$ .

Poslednja osobina znači da ne samo da se dokumenti mogu predstaviti linearom kombinacijom ključnih reči, već se i ključne reči mogu predstaviti linearom kombinacijom dokumenata. Bazu vektorskog prostora, prema tome, ne moraju da čine samo ključne reči, već u njoj mogu biti i dokumenti.

### 3.3 Skalarni proizvod vektora

Neka su dati vektorski prostor  $\Omega$  s vektorima  $\mathbf{x}$ ,  $\mathbf{y}$  i  $\mathbf{z}$  i polje  $\mathbb{F}$  sa skalarom  $\alpha$ . Preslikavanje  $\pi : \Omega \times \Omega \mapsto \mathbb{F}$  za koje važi

$$\pi(\mathbf{x} + \mathbf{y}, \mathbf{z}) = \pi(\mathbf{x}, \mathbf{z}) + \pi(\mathbf{y}, \mathbf{z}), \quad (3.14)$$

$$\pi(\alpha \mathbf{x}, \mathbf{y}) = \alpha \pi(\mathbf{x}, \mathbf{y}), \quad (3.15)$$

$$\pi(\mathbf{x}, \mathbf{y}) = \pi(\mathbf{y}, \mathbf{x}), \quad (3.16)$$

$$\mathbf{x} \neq \mathbf{0} \Rightarrow \pi(\mathbf{x}, \mathbf{x}) > 0 \quad (3.17)$$

naziva se skalarnim proizvodom vektora. U vektorskom prostoru nad poljem realnih brojeva, uobičajeno je da se skalarni proizvod definiše kao

$$\pi(\mathbf{x}, \mathbf{y}) = \mathbf{x} \cdot \mathbf{y} = |\mathbf{x}| |\mathbf{y}| \cos \theta, \quad (3.18)$$

gde je simbolom  $|\cdot|$  označen moduo vektora, dok je  $\theta$  ugao između vektora  $\mathbf{x}$  i  $\mathbf{y}$ . Lako se uočava da važi

$$\mathbf{x} \cdot \mathbf{x} = |\mathbf{x}|^2. \quad (3.19)$$

Vektorski prostor nad kojim je definisan skalarni proizvod naziva se euklidskim prostorom. Za dva vektora u euklidskom prostoru kaže se da su ortogonalni ako je njihov skalarni proizvod jednak nuli. Ako su vektori koji čine bazu prostora uzajamno normalni i normirani (jediničnog modula), tada se takav prostor naziva ortonormalnim.

### 3.4 Primena VSM za pronalaženje informacija

U klasičnim primenama VSM za pronalaženje informacija, prepostavlja se da je broj ključnih reči jednak dimenziji prostora  $\Omega$  [Sal75, Rag86]. Vektori koji odgovaraju ključnim rečima su normirani. Skup ključnih reči, baš kao i skup dokumenata, formira bazu prostora, ali se ne prepostavlja njihova ortogonalnost.

Korisnički upit se predstavlja vektorom  $\mathbf{q}$ ,

$$\mathbf{q} = \sum_{i=1}^n q_i \mathbf{t}_i. \quad (3.20)$$

Dokument  $D_r$  predstavlja se vektorom  $\mathbf{D}_r \in \Omega$  prema (3.11). Za određivanje vrednosti njegovih koordinata, koristi se tzv. TF-IDF metoda (*Term Frequency – Inverse Document Frequency*). Težina pridružena  $i$ -toj ključnoj reči u dokumentu  $D_r$ ,  $d_{ir}$ , računa se kao proizvod frekvencije pojavljivanja te ključne reči u posmatranom dokumentu,  $tf_{ir}$  i tzv. inverzne frekvencije dokumenta,  $idf_i$ ,

$$d_{ir} = tf_{ir} idf_i. \quad (3.21)$$

U literaturi su predložene različite formulacije ovih parametara. Frekvencija pojavljivanja ključne reči najčešće se računa po obrascu

$$tf_{ir} = \frac{N_{ir}}{\sum_{j=1}^n N_{jr}}, \quad (3.22)$$

gde je s  $N_{jr}$  označen broj pojavljivanja  $j$ -te ključne reči u posmatranom dokumentu, mada se negde, kao u npr. [Ram03] izjednačava samo s  $N_{ir}$ .

Za izračunavanje vrednosti parametra  $idf_i$  takođe su predloženi različiti obrasci. Označimo s  $|D|$  ukupan broj dokumenata u prostoru  $\Omega$  i s  $|D \ni d_i|$  broj dokumenata u kojima se pojavljuje ključna reč  $d_i$ . Spärck Jonesova je u radu [Spä72] u kome se prvi put pominje inverzna frekvencija dokumenta predložila obrazac

$$idf_i = \log |D| - \log |D \ni d_i| + 1. \quad (3.23)$$

Salton, Wong i Yang tri godine kasnije predlažu sledeći obrazac [Sal75]:

$$idf_i = \lceil \log |D| \rceil - \lceil \log |D \ni d_i| \rceil + 1, \quad (3.24)$$

dok npr. Lee, Chuang i Seamons koriste varijantu [Lee97]

$$idf_i = \log \frac{|D|}{|D \ni d_i|}. \quad (3.25)$$

Primetimo da je frekvencija pojavljivanja ključne reči lokalni parametar, jer zavisi samo od sadržaja posmatranog dokumenta, dok je inverzna frekvencija dokumenta globalni parametar, jer zavisi od sadržaja svih dokumenata u prostoru  $\Omega$ .

U literaturi su opisane i brojne varijante TF-IDF metode, kao što je npr. BD-ACI-BCA [Zob98, Kem02], u kojoj se na sličan način računa i vektor upita.

**Primer 3.1.** Metodom TF-IDF, odredimo vektor dokumenta *kreditna kartica*. U kolekciji se nalazi 5000 dokumenata, a ključne reči *kredit* i *kartica* javljaju se redom u njih 100 i 500.

*Rešenje.* Primetimo da se svaka od ključnih reči javlja tačno jednom u zadatom dokumentu. Član  $tf$  stoga, prema (3.22) u obe ključne reči ima vrednost  $tf = \frac{1}{1+1} = 0,5$ . Prema (3.25), vrednost člana  $idf$  za ključnu reč *kredit* je  $idf = \log \frac{5000}{100} = 1,7$ , a za ključnu reč *kartica* je  $idf = \log \frac{5000}{500} = 1$ . To znači da se ključnoj reči *kredit* pridružuje vrednost koordinate 0,85, a ključnoj reči *kartica* 0,5. Traženi vektor stoga je (0,85, 0,5). ■

Kao mera sličnosti između dokumenta  $\mathbf{D}_r$  i upita  $\mathbf{q}$  koristi se njihov skalarni proizvod,

$$\text{Sim}(\mathbf{D}_r, \mathbf{q}) = \mathbf{D}_r \cdot \mathbf{q} = \sum_{i=1}^n \sum_{j=1}^n d_{ir} q_j \mathbf{t}_i \cdot \mathbf{t}_j. \quad (3.26)$$

Dокумент koji ima najveću sličnost sa upitom potom se nudi korisniku. Moguć je i alternativni pristup, po kome se korisniku nude svi dokumenti čija je sličnost sa upitom veća od unapred zadatog praga; korisnik tada može preduzeti reakciju u smislu „filtriranja“ skupa ponuđenih dokumenata. Reakcija se izvodi modifikovanjem originalnog upita  $\mathbf{q}$  i opisana je obrascem

$$\mathbf{q}' = \alpha \mathbf{q} + \beta \overline{\{\mathbf{x}_{rel}\}} - \varepsilon \overline{\{\mathbf{x}_{irel}\}}, \quad (3.27)$$

gde je  $\mathbf{q}'$  modifikovani upit,  $\{\mathbf{x}_{rel}\}$  skup rezultata koje korisnik ocenjuje kao relevantne,  $\{\mathbf{x}_{irel}\}$  skup rezultata koje korisnik ocenjuje kao irrelevantne,  $\overline{(\cdot)}$  operator usrednjavanja, a  $\alpha$ ,  $\beta$  i  $\varepsilon$  su nenegativne konstante, najčešće takve da je  $\alpha + \beta + \varepsilon = 1$ . Suština ovakvog modifikovanja upita leži u motivaciji da se originalni upit „pomeri” ka klasteru relevantnih dokumenata.

Ovakav model uspešno je korišten za pretraživanje tekstualnih dokumenata, npr. stručnih članaka ili bibliotečkih jedinica, pri čemu je vektor dokumenta formiran na osnovu Univerzalne decimalne klasifikacije (UDC) ili ključnih reči. Uz minimalne modifikacije, on se može koristiti i za pretraživanje multimedijalnih dokumenata.

**Primer 3.2.** TV3P (*TV Program Personalization for Personal Digital Recorder*) je sistem koji korisnicima omogućava personalizaciju servisa digitalne televizije [Yu04]. Raspoloživi programi opisani su odlikama (atributima) i to naslovom (T), žanrom (G), glumcem (A), ključnom rečju (K) i trajanjem (T). Svaki upit u osnovi predstavlja matricu dimenzija  $m \times 2$ ,

$$\mathbf{P} = \begin{bmatrix} t_1 & w_1 \\ t_2 & w_2 \\ \vdots & \vdots \\ t_m & w_m \end{bmatrix},$$

gde su s  $t_i$ ,  $i = 1, \dots, m$  označene uzajamno ortogonalne odlike, a s  $w_i$  njihove respondentne „težine”, tj. relativni značaji. Odlike u matrici upita poredane su po nerastućem redosledu težina, tj. važi  $(\forall i \in \{i, \dots, m\}) w_i \geq w_{i+1}$ . Radi smanjenja računske složenosti, ovakav upit potom se redukuje na  $n$  najznačajnijih odlika,  $n \leq m$  i predstavlja kao vektor  $\mathbf{P} = (w_1, w_2, \dots, w_n)$ .

Svaki raspoloživi program predstavlja se vektorom dimenzije  $n$ ,  $\mathbf{C} = (u_1, u_2, \dots, u_n)$ , čije koordinate odgovaraju težinama pojedinih odlika. Težine odlika programa računaju se po sledećem algoritmu:

- Nominalne težine naslova, žanra, glumca i ključne reči programa, redom su  $w_T = 1,25$ ,  $w_G = 1$ ,  $w_A = 0,75$  i  $w_K = 0,5$ .
- Ako se odlika  $t$  pojavljuje samo u jednom od polja {naslov, žanr, glumac, ključna reč}, tada je njena težina jednaka nominalnoj težini tog polja.
- Ako se odlika pojavljuje u više polja, tada je njena težina jednaka najvećoj težini polja u kojima se pojavljuje.
- Ako se odlika ne pojavljuje ni u jednom od polja, tada je njena težina jednaka nuli.

Naredni korak predstavlja računanje sličnosti programa  $\mathbf{C}$  i upita  $\mathbf{P}$ , za šta se koristi kosinus ugla između njihovih vektora,

$$\text{Sim}(\mathbf{C}, \mathbf{P}) = \frac{\mathbf{C} \cdot \mathbf{P}}{|\mathbf{C}| |\mathbf{P}|} = \frac{\sum_{i=1}^n u_i w_i}{\sqrt{\sum_{i=1}^n u_i^2} \sqrt{\sum_{i=1}^n w_i^2}}.$$

Ako je sličnost posmatranog programa s upitom veća od zadatog praga (0,40), smatra se da je program relevantan za tekući upit i on se nudi korisniku.

Ispitajmo je li film „Prohujalo s vihorom”, čiji je opis dat sledećim odlikama:

Naslov: Prohujalo s vihorom

Žanr: Romansa

Glumci: Clark Gable, Vivien Leigh...

Ključne reči: Ljubav, brak, rat...

Trajanje: 180 minuta.

relevantan za originalni upit

$$\mathbf{P} = \begin{bmatrix} \text{romansa} & 1,25 \\ \text{animacija} & 1,16 \\ \text{ljubav} & 1,10 \\ \text{akcija} & 1,00 \\ \text{brak} & 0,84 \\ \text{sport} & 0,78 \\ \text{religija} & 0,72 \\ \text{oružje} & 0,60 \\ \vdots & \vdots \end{bmatrix}.$$

*Rešenje.* Redukovani upit se dobija tako što se iz originalnog zadrži npr. prvih šest odlika:

$$\mathbf{P} = \begin{bmatrix} \text{romansa} & 1,25 \\ \text{animacija} & 1,16 \\ \text{ljubav} & 1,10 \\ \text{akcija} & 1,00 \\ \text{brak} & 0,84 \\ \text{sport} & 0,78 \end{bmatrix}$$

i potom se posmatraju samo njihove težine,

$$\mathbf{P} = (1,25, 1,16, 1,10, 1,00, 0,84, 0,78).$$

Posmatrajmo sada koje se odlike iz redukovanog upita pojavljuju u opisu programa. Odlika *romansa* pojavljuje se jedanput u polju „žanr”, pa ima težinu koja je jednak nominalnoj, tj. 1. Odlike *ljubav* i *brak* pojavljuju se takođe jednom i to u polju „ključna reč”, pa su im težine po 0,5. Odlike *animacija*, *akcija* i *sport* ne pojavljuju se u opisu filma „Prohujalo s vihorom”, pa su im težine jednake nuli. To znači da je vektor programa

$$\mathbf{C} = (1, 0, 0, 5, 0, 0, 5, 0).$$

Konačno dobijamo da je sličnost filma „Prohujalo s vihorom” s pretpostavljenim upitom  $\text{Sim}(\mathbf{C}, \mathbf{P}) = 0,71$ . Pošto je sličnost veća od praga, ovaj film je relevantan za upit i nudi se korisniku. ■

### 3.5 Modeliranje multimedijalnog dokumenta

Dva važna pitanja koja se postavljaju pri razmatranju mogućnosti primene modela vektorskog prostora za pronalaženje multimedijalnih dokumenata i servisa su:

- (1) Kako izabrati bazu vektorskog prostora i
- (2) Kako, za zadatu bazu, predstaviti multimedijalni dokument ili telekomunikacioni servis vektorom.

Ova pitanja nisu nezavisna i na njih, nažalost, još uvek nema potpunog i opšteg odgovora. Primetimo da prvo pitanje zadire u samu suštinu percepcije multimedijalnog sadržaja. Može se očekivati da bi imalo smisla da bazu vektorskog prostora čine upravo oni elementi koji su ključni za doživljavanje multimedije, ali saznanja o tom procesu i dalje su oskudna. Klasična znanja iz fiziologije koja imamo o funkcijanju ljudskih čula ne mogu se pravolinijski ekstrapolirati na percepciju multimedije. Iako je dobro poznato kako doživljavamo pojedinačne draži, još uvek ne znamo mnogo o percepciji njihovog združenog dejstva. Psihometrijska istraživanja i dalje se sprovode na malim uzorcima ispitanika (10-15, [Ito05]), pa je upitno koliko su opšti njihovi rezultati i zaključci. Indeksiranje multimedijalnih sadržaja zbog toga može da postane isuviše deskriptivno — npr. film bi se mogao opisati tehničkim detaljima (trajanje, format, kolor, zvuk), sadržajem (žanr, podžanr), imenima glumaca, reditelja, podacima o nagradama itd. Očigledno, ovakav pristup bi doveo do vektorskog prostora nedopustivo velikih dimenzija. Na pitanju modeliranja multimedije stoga se intenzivno radi, za šta dobar primer pruža rad [Vol07].

Tabela 3.1: Primer *metadate*.

Polje	Vrednost
<Program Code>	EP1151270151
<Title>	Friends
<Short Title>	Friends
<Episode Title>	The One With the Wedding
<Synopsis>	Rachel serves as a bridesmaid in her ex-fiancé Barry's wedding.
<Genre>	Situation, Comedy
<Channel>	WPIX
<Air Time>	2000)
<Air Date>	20020912
<Actors>	Jennifer Aniston, Courtney Cox, Lisa Kudrow, Matt LeBlanc, Matthew Perry, David Schwimmer
<Producers>	Marta Kauffman
<Language>	English

Moguće rešenje ovog problema već sada se nazire iz radova na standardizaciji u multimediji, naročito na standardima MPEG-7 i MPEG-21. Naime, audiovizuelni sadržaji mogu se indeksirati korišćenjem sprege za opis multimedijalnog sadržaja MPEG-7 [Tse04]. MPEG-7 pruža standardizovane opise za pretraživanje, filtriranje, izbor i

obradu audiovizuelnog sadržaja. Ovi opisi, koji se nazivaju *metadata*, mogu se odnositi na (1) upravljanje audiovizuelnim podacima i (2) na određene specifične koncepte i odlike. U prvom slučaju, metadata opisuje stvaranje, produkciju, korišćenje sadržaja i upravljanje njime, a u drugom opisuje npr. o čemu se radi u nekoj sceni video-klipa, koji su objekti prisutni, ko govori, kakva je raspodela boja itd.

U tabeli 3.1, dat je primer *metadata* opisa epizode jedne televizijske serije [Zim04], koji pripada slučaju (1). Da bi se omogućio opis za slučaj (2), MPEG-7 definiše simbolički jezik i postupke opisivanja različitih nivoa apstrakcije, varijacije i semantičkog značenja multimedijalnog sadržaja. Korišćenjem *metadata* informacije, na ovaj način se svakom multimedijalnom dokumentu može inherentno pridružiti njegov opis, čime se omogućava jednoznačno predstavljanje tog dokumenta u modelu vektorskog prostora. Značaj *metadate* za personalizaciju telekomunikacionih servisa ovde se ne završava, jer se informacija iz nje može kasnije iskoristiti za modeliranje korisnika koji interaguje s posmatranim multimedijalnim dokumentom [Kay04].

## 3.6 Relevantnost dokumenta za upit

Neka su dati vektorski prostor  $\Omega$  i vektori  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ ,  $\mathbf{y} = (y_1, y_2, \dots, y_n)$  i  $\mathbf{z} = (z_1, z_2, \dots, z_n)$ . Pitanje koje se postavlja je kako iskazati „sličnost” ili „različitost” ovih vektora, pri čemu se podrazumeva da oni opisuju multimedijalne sadržaje, koje opažamo po nelinearnim zakonitostima.

U prvobitnoj formulaciji modela vektorskog prostora, kao mera sličnosti dvaju dokumenata korišten je skalarni proizvod njihovih vektora, (3.26), koji se može interpretirati kao korelacija između ovih vektora. Zbog specifičnosti multimedijalnog sadržaja, koriste se i druge funkcije, koje su prikazane u nastavku ovog odeljka; u najgrubljoj podeli, možemo ih svrstati u dve grupe, mere razdaljine i mere sličnosti.

### 3.6.1 Mere razdaljine

Najpoznatije mere razdaljine su metričke funkcije. Neka je data funkcija  $f$ , takva da važi

$$f(\mathbf{x}, \mathbf{y}) \geq 0, \quad (3.28)$$

$$f(\mathbf{x}, \mathbf{y}) = 0 \iff \mathbf{x} = \mathbf{y}, \quad (3.29)$$

$$f(\mathbf{x}, \mathbf{y}) = f(\mathbf{y}, \mathbf{x}), \quad (3.30)$$

$$f(\mathbf{x}, \mathbf{y}) + f(\mathbf{y}, \mathbf{z}) \geq f(\mathbf{x}, \mathbf{z}). \quad (3.31)$$

Funkcija  $f$  naziva se *metrikom* i ima interpretaciju rastojanja između svojih argumenta – što su dva vektora „sličniji”, njihovo rastojanje u posmatranom prostoru je manje i obrnuto.

U nekim primenama su od interesa i funkcije koje ne zadovoljavaju sve uslove (3.28)-(3.31). Ako nije ispunjen uslov (3.29), funkcija  $f$  naziva se *pseudometrikom*. Ako ne važe uslovi (3.29) i (3.30), radi se o *hemimetrici*. Ako ne važi samo (3.30), funkcija  $f$  je *kvazimetrika*. Za funkciju  $f$  kaže se da je *semimetrika* ako ne zadovoljava samo nejednakost trougla, (3.31). Konačno, ako je zadovoljen samo uslov (3.28) uz  $f(\mathbf{x}, \mathbf{x}) = 0$ , funkcija  $f$  naziva se *prametrikom*.

Prva mera razdaljine koju ćemo razmotriti je metrika Minkowskog, koja je data izrazom

$$f(\mathbf{x}, \mathbf{y}) = \sqrt[p]{\sum_{i=1}^n |x_i - y_i|^p}, \quad (3.32)$$

pri čemu je parametar  $p$  realan broj koji nije manji od 1.

Metrika Minkowskog ima tri važna posebna slučaja. Za  $p = 1$ , dobija se Manhattan ili *taxi cab* metrika,

$$f(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n |x_i - y_i|. \quad (3.33)$$

Ova metrika je dobila naziv po tome što odgovara najkraćem rastojanju koje u gradskim centrima treba preći da bi se došlo od tačke  $\mathbf{x}$  do tačke  $\mathbf{y}$ .

Za  $p = 2$ , radi se o Euklidovoj metrići:

$$f(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}. \quad (3.34)$$

Kada  $p \rightarrow \infty$ , dobija se metrika supremuma,

$$f(\mathbf{x}, \mathbf{y}) = \max_i |x_i - y_i|. \quad (3.35)$$

Metrici Minkowskog slična je  $l_p$  metrika,

$$f(\mathbf{x}, \mathbf{y}) = \sqrt[p]{\frac{1}{n} \sum_{i=1}^n |x_i - y_i|^p}, \quad (3.36)$$

pri čemu je  $p$  sada prirodan broj. Normalizacioni član  $1/n$  u literaturi se ponekad izostavlja.

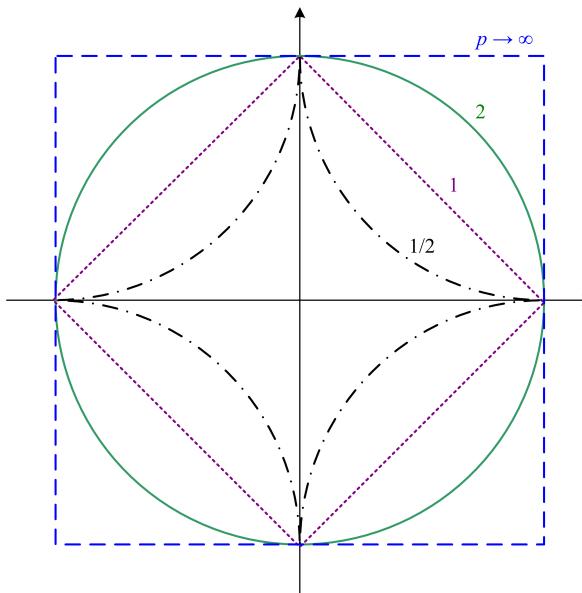
Ponderisana varijanta Euklidove metrike naziva se Mahalanobisovim rastojanjem,

$$f(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^n w_i (x_i - y_i)^2}, \quad (3.37)$$

gde su sa  $w_i$ ,  $i = 1, \dots, n$  označeni težinski koeficijenti. Izborom njihovih vrednosti uzima se u obzir nejednak uticaj pojedinih koordinata vektorskog prostora na percepciju servisa.

S povećavanjem dimenzije vektorskog prostora,  $n$ , performanse metrike Minkowskog, u smislu sposobnosti distinkcije bliskih vektora, rastu sa smanjivanjem vrednosti parametra  $p$  [Agg01]; stoga ima smisla da se u prostorima velikih dimenzija koristi tzv. frakcionala metrika, koja se dobija kada se u (3.32) dozvoli da vrednost parametra  $p$  pripada intervalu  $(0, 1)$ .

Na slici 3.1, prikazane su izodistante linije funkcije (3.32), za različite vrednosti parametra  $p$ . Radi preglednosti, posmatran je dvodimenzionalni prostor, pri čemu je prepostavljeno da se jedan od vektora nalazi u koordinatnom početku.



Slika 3.1: Primeri izodistantnih linija funkcije (3.32), za različite vrednosti parametra  $p$ , u dvodimenzionalnom vektorskem prostoru.

Zeta rastojanje [Che05] dato je izrazom

$$f(\mathbf{x}, \mathbf{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i^2 - y_i^2)^2}. \quad (3.38)$$

Ova funkcija je primer pseudometrike. Ona postaje metrika onda kada su sve koordinate vektora  $\mathbf{x}$  i  $\mathbf{y}$  istog znaka.

U radu [Gót05], predložena je upotreba hiperboličkog rastojanja,

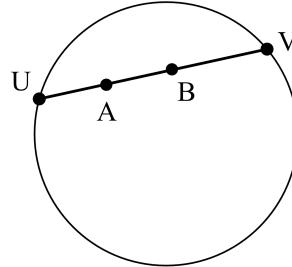
$$f(\mathbf{x}, \mathbf{y}) = \left| \ln \frac{r - \sqrt{\sum_{i=1}^n (x_i - y_i)^2}}{r + \sqrt{\sum_{i=1}^n (x_i - y_i)^2}} \right|, \quad (3.39)$$

pri čemu je  $r$  podešljivi parametar. Geometrijska interpretacija ovoga rastojanja u ravni bila bi sledeća. Posmatrani hiperbolički prostor predstavlja kružnu površ. Neka

su date tačke A i B koje pripadaju ovome prostoru. Neka je, dalje, kroz ove tačke konstruisana tetiva kruga, čije su krajnje tačke U i V, kao što je to prikazano na slici 3.2. Hiperboličko rastojanje tačaka A i B tada je

$$f(A, B) = \left| \ln \frac{d_E(A, U) d_E(B, V)}{d_E(A, V) d_E(B, U)} \right|, \quad (3.40)$$

gde je  $|\cdot|$  operator absolutne vrednosti i  $d_E$  Euklidovo rastojanje račaka.



Slika 3.2: Geometrijska interpretacija hiperboličkog rastojanja (3.39) u ravni.

Ovaj pregled funkcija razdaljine završićemo tzv. merama divergencije, koje se definišu nad prostorima koji su ograničeni na pozitivni hiperkvadrant [Kos07].

Kullback-Leiblerova divergencija data je izrazom

$$f(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n x_i \log \frac{x_i}{y_i}. \quad (3.41)$$

Ova funkcija ima smisla samo ako nijedna koordinata vektora  $\mathbf{y}$  nije jednaka nuli. Primetimo da ona nije ni prametrika, jer ne zadovoljava uslov (3.28); takođe, nije zadovoljen ni uslov simetrije (3.30). Definiše se i simetrična varijanta Kullback-Leiblerove divergencije, koja ga zadovoljava:

$$f(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n x_i \log \frac{x_i}{y_i} + y_i \log \frac{y_i}{x_i}. \quad (3.42)$$

Jeffreyjeva divergencija data je izrazom

$$f(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n x_i \log \frac{x_i}{\xi_i} + y_i \log \frac{y_i}{\xi_i}, \quad (3.43)$$

gde je  $\xi_i = (x_i + y_i) / 2$ ,  $i = 1, \dots, n$ .

### 3.6.2 Mere sličnosti

Funkcije sličnosti nemaju formalnu matematičku definiciju; za pronalaženje informacija bilo bi poželjno da zadovoljavaju sledeće uslove:

$$0 \leq f(\mathbf{x}, \mathbf{y}) \leq 1, \quad (3.44)$$

$$f(\mathbf{x}, \mathbf{y}) = f(\mathbf{y}, \mathbf{x}), \quad (3.45)$$

$$f(\mathbf{x}, \mathbf{x}) = \kappa = \text{const.} \quad (3.46)$$

Najveću primenu ima kosinusna sličnost, s kojom smo se upoznali u primeru 3.2:

$$f(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x} \cdot \mathbf{y}}{|\mathbf{x}| |\mathbf{y}|} = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}}. \quad (3.47)$$

Ova funkcija može se shvatiti kao kosinus ugla između posmatranih vektora, kao njihov normalizovani skalarni proizvod, ili kao korelacija. Ako je prostor  $\Omega$  ograničen na prvi hiperkvadrant, tada kosinusna sličnost zadovoljava uslove (3.44)-(3.46), uz  $\kappa = 1$ .

Pseudokosinusna sličnost data je izrazom

$$f(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}. \quad (3.48)$$

Zanimljivo je da u prostoru ove sličnosti postoje vektori koji nisu najsličniji samim sebi.

Dve naredne funkcije spadaju u grupu tzv. mera asocijacije. Diceov koeficijent definiše se kao

$$f(\mathbf{x}, \mathbf{y}) = \frac{2 \sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i + \sum_{i=1}^n y_i}. \quad (3.49)$$

Koeficijent preklapanja, koji je korišten u prvim sistemima za automatsko pronalaženje informacija, definisan je izrazom

$$f(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^n \min(x_i, y_i)}{\min\left(\sum_{j=1}^n x_j, \sum_{k=1}^n y_k\right)}. \quad (3.50)$$

Ako je data (pseudo)metrička funkcija  $\mu$ , tada se iz nje može izvesti funkcija sličnosti  $f$  koja zadovoljava uslove (3.44)-(3.46) na sledeći način:

- (1) Funkcija  $\mu$  normalizuje se na posmatranom prostoru  $\Omega$ , tako da važi

$$(\forall \mathbf{x}, \mathbf{y} \in \Omega) \quad \mu_n(\mathbf{x}, \mathbf{y}) \in (0, 1), \quad (3.51)$$

(2) Formira se funkcija

$$f(\mathbf{x}, \mathbf{y}) = 1 - \mu_n(\mathbf{x}, \mathbf{y}), \quad (3.52)$$

koja predstavlja traženu funkciju sličnosti.

Na ovaj način, polazeći od hiperboličkog rastojanja, (3.39), može se izvesti hiperbolička sličnost

$$f(\mathbf{x}, \mathbf{y}) = \left( \ln \left( e \frac{r + \sqrt{\sum_{i=1}^n (x_i - y_i)^2}}{r - \sqrt{\sum_{i=1}^n (x_i - y_i)^2}} \right) \right)^{-1}. \quad (3.53)$$

## 3.7 Pitanja i zadaci

**3.1.** Objasnite prirodu i značaj frekvencije ključne reči i inverzne frekvencije dokumenta.

**3.2.** Kolekcija se sastoji od dokumenata  $D_1$ ,  $D_2$  i  $D_3$ , čije su karakteristike date u tabeli ( $n$  je broj pojavljivanja ključne reči u dokumentu, a  $K$  broj dokumenata koji sadrže ključnu reč).

Pojam	$n_{D_1}$	$n_{D_2}$	$n_{D_3}$	$K$
actor	12	35	55	123
movie	15	24	48	240
trailer	52	13	12	85

(a) Odredite vektore dokumenata  $D_1$ ,  $D_2$  i  $D_3$  korišćenjem jedne od varijanti metode TF-IDF.

(b) Odredite uzajamne kosinusne sličnosti ovih dokumenata.

(c) Odredite rangiranje (poredak) dokumenata za upit `movie trailer`.

**3.3.** Objasnite zbog čega Jaccardovo rastojanje,

$$d_J(D, Q) = 1 - \frac{|D \cap Q|}{|D \cup Q|},$$

nije pogodno za primenu kao funkcija rangiranja u sistemima za pronalaženje informacija: konstruišite dokument  $D_1$  koji će biti relevantan za upit  $Q = \text{open source}$  i dokument  $D_2$  koji neće, a za koje će važiti  $d_J(D_1, Q) > d_J(D_2, Q)$ .

**3.4.** Objasnite zbog čega recipročna vrednost Euklidovog rastojanja nije pogodna kao mera sličnosti u vektorskom prostoru.

**3.5.** Konstruišite vektor koji u prostoru pseudokosinusne sličnoati (3.48) nije najsličniji samome sebi.

**3.6.** Polazeći od izraza za hiperboličko rastojanje (3.39), izvedite izraz za hiperboličku sličnost (3.53).

**3.7.** Konstruišite izosimilarne linije hiperboličke sličnosti.

**3.8.** Pokažite da hiperbolička sličnost daje isti poredak dokumenata prema upitu kao i kosinusna.



# Poglavlje 4

## Modeliranje korisnika

### 4.1 Uvod

Cilj modeliranja korisnika je prilagođavanje ponašanja sistema kao celine potrebama korisnika kao individue. Počeci modeliranja korisnika vezuju se za kraj sedamdesetih godina prošloga veka [Wah86, Kob01]. S izuzetkom sistema razvijenih za čisto akademske potrebe, koji su, na primer, „razgovarali“ s gostima na prijemu, prve praktične primene modeliranja korisnika bile su u sistemima za pružanje informacija na železničkim stanicama, obračun poreza, rezervisanje hotelskih soba i preporučivanje knjiga u biblioteci. Modeliranje korisnika danas se široko koristi u primenama kao što su pretraživanje i filtriranje informacija, elektronsko učenje, adaptivni korisnički interfejsi, „inteligentni“ korisnički agenti itd.

U nastavku ove glave, u prvom delu su opisane postavke modeliranja korisnika koje se zasniva na tehnikama mašinskog učenja. Naročita pažnja posvećena je odlikama koje su važne za personalizaciju telekomunikacionih servisa. U drugom delu glave, pokazano je kako se modeliranje korisnika može izvesti u modelu vektorskih prostora.

### 4.2 Sistemi za modeliranje korisnika

Mehanizmi za modeliranje korisnika treba da sistemima koji prilagođavaju svoj rad karakteristikama pojedinačnog korisnika obezbede eksplicitan skup informacija o preferencijama posmatranog korisnika, njegovim (ili njenim) interesovanjima, ciljevima, znanju itd. [Poh96]. Ovakav skup informacija naziva se modelom korisnika.

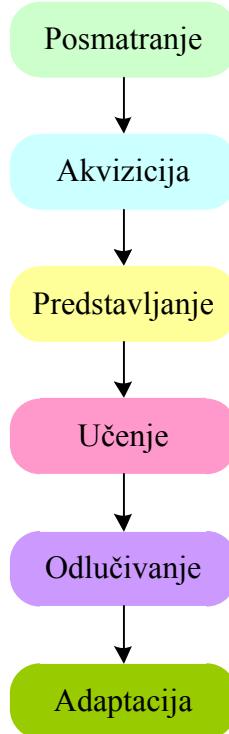
Osnovni servisi koje nude savremeni sistemi za modeliranje korisnika su [Kob01]:

- predstavljanje prepostavki o odlikama pojedinačnih korisnika u njihovim mode-

lima,

- predstavljanje relevantnih zajedničkih odlika više korisnika koji pripadaju istoj podgrupi korisnika aplikacionog sistema u cilju formiranja tzv. stereotipa,
- klasifikacija korisnika koji pripadaju ovim podgrupama i integracija tipičnih odlika podgrupe u tekuće pojedinačne korisničke modele,
- praćenje interakcija korisnika sa sistemom,
- izvođenje pretpostavki o korisniku na osnovu podataka o prethodnim interakcijama,
- generalizacija toka interakcija više korisnika u stereotipe,
- izvođenje dodatnih pretpostavki o korisniku na osnovu početnih,
- održavanje konzistentnosti modela korisnika,
- obezbeđivanje tekućih pretpostavki o korisniku, kao i opravdanja za njih i
- poređenje modela korisnika sa zadatim standardima.

Procesi koji se odvijaju pri modeliranju korisnika prikazani su na slici 4.1.



Slika 4.1: Tok procesa u sistemu za modeliranje korisnika – prema [Poh99], modifиковано.

Sistem posmatra aktivnosti korisnika, registruje ih i predstavlja u pogodnoj formi. Iz svake nove aktivnosti, sistem „uči” i formira model korisnika. Na osnovu znanja sadržanog u modelu korisnika, sistem donosi odluke u smislu adaptacije ponašanja u skladu s anticipiranim potrebama posmatranog korisnika.

Iz ovoga prikaza vidimo da je centralni element u procesu prilagođavanja sistema potrebama korisnika upravo model korisnika. Formalno definisan, model korisnika predstavlja izvor znanja koji sadrži eksplicitne pretpostavke o svim aspektima posmatranog korisnika koje mogu biti relevantne za ponašanje sistema [Wah86].

Dva važna pitanja koja se sada postavljaju su [Pap01]:

- (1) Koji parametri su relevantni za model korisnika i
- (2) Kako, za usvojenu formu modela, izvršiti njegovu akviziciju.

Odgovor na prvo pitanje zavisiće od toga šta je cilj opisivanja korisnika, odnosno, šta posmatrani sistem treba da obavi. Model korisnika tako može da opisuje:

- kognitivne procese na kojima se zasnivaju aktivnosti korisnika,
- razlike u sposobnostima posmatranog korisnika i eksperta za datu oblast,
- obrasce ili preferencije ponašanja korisnika ili
- opšte karakteristike korisnika.

U početku, sistemi za modeliranje korisnika su se naročito koncentrisali na prvi i nešto manje na drugi tip modela. Nasuprot tome, u savremenim sistemima primenjen je treći pristup, a radi se na razvoju četvrtog [Web01].

U savremenim sistemima, model korisnika se iskazuje u vidu skupa preferencija posmatranog korisnika, pravila i stabala odlučivanja, a može sadržavati i lične podatke o korisniku, kao što su npr. uzrast i zanimanje. Sa stanovišta primene za personalizaciju telekomunikacionih servisa, model se najčešće predstavlja skupom uređenih parova (*parametar, vrednost*), što podseća na model vektorskog prostora, ili u vidu relacionih baza podataka. Detaljan pregled različitih varijanti modela korisnika dat je u [Bou06].

Razmotrimo sada pitanje akvizicije modela korisnika. Za to postoje barem četiri načina [Poh96, Ard04] i to:

- eksplicitno isticanje interesovanja i želja od strane korisnika, na primer putem intervjua ili popunjavanjem upitnika,
- korišćenje socio-demografskih informacija o korisniku, kao što su, na primer, starost, školska spremna, zanimanje,
- korišćenje informacija o životnom stilu korisnika (npr. opšta interesovanja, hobiji),
- korišćenje informacija izvedenih iz prethodnih aktivnosti korisnika.

U nastavku ćemo razmotriti svaki od njih.

Očigledno je da eksplicitno isticanje preferencija dovodi do formiranja modela koji na najbolji način opisuje posmatranog korisnika. Eksplicitno modeliranje, međutim, ne sme biti samo po sebi cilj, jer prilikom njega sistem preuzima kontrolu nad interakcijom, a korisnik se stavlja u drugi plan. Sigurno većina korisnika interneta iz ličnog iskustva može posvedočiti koliko je neugodno i zamorno popunjavanje različitih upitnika koje, najčešće pod formom registracije, prethodi pristupu nekim sadržajima. Eksplicitno modeliranje stoga treba koristiti samo u izuzetnim slučajevima, najbolje na zahtev korisnika, ili kada sistem zaključi da postojeći model značajno odstupa od trenutnih aktivnosti korisnika.

Korišćenjem socio-demografskih informacija i informacija o životnom stilu, korisnik se može svrstati u neku od predefinisanih stereotipnih grupa s poznatim preferencijama

u pogledu pristupanja posmatranom telekomunikacionom servisu ili multimedijalnom sadržaju. Preferencije za svaku od stereotipnih grupa izvode se na osnovu opsežnih socio-oloških i psihometrijskih istraživanja. Na slici 4.2, dat je primer stereotipa „Domaćica“ u jednom sistemu za preporučivanje televizijskih programa.

Domaćica				
	Vrednost	Incidenca	Parametar	Značaj
Uzrast	< 15	0	Ljubavni film	1
	15 – 24	0	Vesti	0,2
	25 – 34	0	Moda	0,5
	35 – 44	0,5	Kuvanje	1
	45 – 54	0,5	Serije	1
	55 – 64	0		
	> 64	0		
Pol	Muški	0		
	Ženski	1		

Slika 4.2: Primer stereotipa [Ard04].

Sa slike se vidi da se prikazani stereotip sastoji od dveju grupa podataka, koje su organizovane tabelarno. Levu tabelu čine socio-demografske informacije, a desnou podaci od neposrednog interesa za formiranje modela korisnika. Prva grupa podataka služi kao reper na osnovu koga se procenjuje može li se korisnik svrstati u posmatrani stereotip; statistički testovi predstavljaju objektivno merilo podudaranja za to. Ako je rezultat testa pozitivan, korisnik se podvodi pod posmatrani stereotip i njegovom modelu se pripisuju vrednosti parametara koje se, u primeru sa slike 4.2, nalaze u desnoj tabeli. Ako sistem već raspolaže vrednostima nekih parametara modela toga korisnika, tada se umesto potpunog prepisivanja iz stereotipnog modela mogu uzeti samo nedostajuće vrednosti. Postoje i stereotipi koji umesto socio-demografskih informacija sadrže informacije o životnom stilu korisnika; u tome slučaju kaže se da se radi o svrstavanju korisnika u zajednicu.

Korišćenjem stereotipa se neposredno posle registracije korisnika, a pre prvog pristupanja servisu, formira inicijalni model korisnika. Ovim se rešava tzv. problem hladnog starta, tj. kako u odsustvu bogatog ličnog profila novim korisnicima ponuditi sadržaje koji bi ih interesovali. U većini postojećih sistema i aplikacija, modeliranje korisnika (i kasnija personalizacija sadržaja) na ovom se mestu i završava. Korisnik tako, na primer, prilikom pristupanja internet-pretraživaču vidi uputstva na svom jeziku, jer sistem na osnovu IP adrese njegovog proksi servera „prepoznaće“ iz koje zemlje dolazi, ili na sajtu s meteorološkim podacima vidi samo one podatke koji se odnose na njegov grad.

Primetimo da je koncept korišćenja stereotipa blizak ideji kolaborativnog filtriranja, po kojoj se adaptacija za konkretnog korisnika izvodi posmatranjem aktivnosti srodnih

korisnika u sistemu. Savremeni pristup modeliranju korisnika podrazumeva korišćenje tehnika mašinskog učenja, tako da se iz interakcija korisnika s nudiočem sadržaja izvode zaključci koji se koriste za fino podešavanje inicijalnog stereotipnog modela. U pomenutom primeru s meteorološkim internet sajtom, to bi značilo da, ako je korisnik prethodno posetio sajt turističke agencije i razgledao aranžmane za Austriju, sistem mu sada nudi dugoročnu prognozu za tu zemlju. Neke jednostavne heuristike koje se mogu koristiti pri učenju modela korisnika i donošenju odluka navedene su u [Poh95]. To su, na primer:

- Ako korisnik pravilno upotrebljava neke objekte (komande na uređaju, funkcionalnosti u računarskom programu), tada se može pretpostaviti da je upoznat s njima.
- Ako, pak, korisnik upotrebljava objekte na pogrešan način, tada nije upoznat s njima.
- Ako korisnik traži objašnjenje o nekom objektu, tada nije upoznat s njime.
- Ako korisnik traži detaljniju informaciju o objektu, tada već poseduje izvesno znanje o njemu, itd.

Opisaćemo sada u najkraćim crtama jedan sistem za modeliranje korisnika opšte namene.

Doppelgänger („dvojnik“ na nemačkom) [Orw95] jedan je od prvih uspešnih sistema za modeliranje korisnika koji koristi tehnike mašinskog učenja. Relevantne informacije o korisniku prikupljaju se hardverskim i softverskim putem, što znači (1) neposrednom akvizicijom iz fizičkog sveta i (2) kasnijom obradom ovakvih informacija, da bi se dobile informacije višeg reda. U zavisnosti od konkretne primene ovog sistema opšte namene, koriste se različite tehnike formiranja modela korisnika, na primer:

- Zaključci se izvode iz binarnih posmatranja korisnikovog ponašanja (npr. korisniku se neki novinski članak sviđa, ili mu se ne sviđa). Stepen poverenja u ovakve zaključke određuje se statističkim testovima.
- Naredni termin u kome će se korisnik prijaviti na sistem (tj. koristiti aplikaciju) određuje se linearnom predikcijom, na osnovu posmatranja njegovih prethodnih sesija.
- Tranzicije u navikama korisnika (npr. rad/odmor) predviđaju se pomoću Markovljevih modela.
- Koristi se algoritam za klasterizovanje modela korisnika u modele zajednica, koje imaju ulogu stereotipa. Nedostajuće informacije u modelu pojedinačnog korisnika zamenuju se podacima iz modela najsličnije zajednice.

Savremeni sistemi za modeliranje korisnika zasnivaju se na arhitekturi klijent-server [Kob01]. Informacije o korisnicima čuvaju se u centralnom repozitorijumu i na raspolaganju su svim aplikacijama sistema. Informacije koje se prikupe tokom izvršavanja jedne aplikacije stavlja se na raspolaganje i svim ostalim aplikacijama u sistemu. Ove informacije mogu se odnositi na modele pojedinačnih korisnika, na apriorne stereotipe, na modele zajednica, ili na modele grupe korisnika. Naročita pažnja se posvećuje problemu zaštite privatnosti korisnika, kroz implementiranje mehanizama kriptozaštite i kontrole pristupa podacima o korisnicima, uključujući i njihove modele.

Na kraju ovoga pregleda, osvrnućemo se na faktore koji trenutno ograničavaju punu primenu tehnika modeliranja korisnika. To su [Web01]:

- potreba za opšenim skupovima podataka za učenje,
- potreba za „označavanjem“ podataka, da bi se znalo jesu li oni relevantni ili irrelevantni za korisnika. Ovo pitanje je povezano s tehnikama reakcije po značaju, koje su razmotrene u narednom odeljku,
- „drift“ koncepta, što znači da se algoritmi za učenje moraju brzo prilagođavati promenama u okruženju i
- složenost algoritma, koja se odražava kroz zahteve za vremenom izvršavanja, procesorskom moći, memorijom, potrošnjom energije itd.

### 4.3 Reakcija po značaju

Na osnovu važećeg modela korisnika, iz skupa raspoloživih servisa/sadržaja biraju se oni koji su za njega potencijalno interesantni. Korisnik može da prihvati neku od ovih ponuda, ili da ih sve odbije. U zavisnosti od toga se, prema usvojenom algoritmu, ažurira važeći model posmatranog korisnika. Ovde prepoznajemo koncept povratne sprege, s tim što se u slučaju modeliranja korisnika govori o povratnoj sprezi ili reakciji po značaju.

Neki sistemi koriste eksplicitnu reakciju, u kojoj se od korisnika traži da oceni važnost ponuđenih sadržaja. Na primer, u slučaju internet-pretraživača, korisniku se nudi nekoliko odgovora na upit i od njega se traži da ih oceni, „naslepo“, na osnovu ponuđenih objašnjenja, ili po izvršenom uvidu u svaki od njih. Ovakav pristup se naziva invazivnim i za njega važe isti komentari kao i za eksplicitno modeliranje.

U slučaju implicitne reakcije, informacija o korisniku se izvodi iz posmatranja njegove prirodne interakcije sa sistemom, pa se za ovakav pristup kaže da je neinvazivan. Moguća ponašanja korisnika na osnovu kojih bi se mogla izvršiti reakcija su [Kel03]:

- uvid u dokument/sadržaj: čitanje, slušanje, gledanje, skrolovanje, pretraživanje unutar dokumenta;
- zadržavanje dokumenta: štampanje, snimanje, slanje elektronskom poštom, pretplaćivanje na sadržaj;
- referenciranje dokumenta: citiranje, prosleđivanje, kopiranje;
- označavanje delova dokumenta;
- pravljenje novog dokumenta itd.

Pored ovih akcija, u nekim sistemima se registruje vreme u koje se koristi aplikacija, broje se posete internet stranicama od značaja, meri se vreme provedeno čitajući elektronsku poštu itd. Očekuje se da bi se implicitna reakcija u budućnosti mogla zasnovati i na biometrijskim podacima – posmatranju izraza lica i jezika tela [Web01].

Iako dobro zvuči, implicitna reakcija u sebi nosi prilične kontroverze. Svaki pokušaj

izvođenja informacija o preferencijama korisnika posmatranjem njegovih aktivnosti povezan je s brojnim problemima, jer osmotrive aktivnosti ne korespondiraju jednoznačno sa stvarnim namerama korisnika. Na primer, mnogi sistemi, a među njima i TV3P [Yu04], koji smo pomenuli u prethodnoj glavi, posmatraju koliko je vremena korisnik proveo gledajući neki sadržaj, jer se očekuje da će duže gledati sadržaje koji ga više interesuju. Pri tome se zanemaruje činjenica da korisnik može, na primer, da pogleda emisiju vesti samo do one informacije koja ga zaista zanima i da potom prekine dalje gledanje. Neki postojeći sistemi bi ovakvu aktivnost korisnika protumačili kao nezainteresovanost za program vesti, što je potpuno pogrešno. Zbog toga se u [Tam05] kategorički navodi da tehnike modeliranja korisnika koje se zasnivaju na implicitnoj reakciji nisu efikasne u stvarnim primenama. Sistemi stoga treba da nađu meru u količini i tipu korišćene reakcije, tako da efikasno prikupe parametre koji su relevantni za opis korisnika, a da pri tome ne budu previše invazivni.

## 4.4 Problem zaštite privatnosti

Pre nego što počnu da se prilagođavaju željama i potrebama korisnika, sistemi za personalizaciju treba da prikupe izvesnu količinu podataka o njemu. Dobre strane personalizacije stoga će moći da dodu do izražaja samo u onim slučajevima kada korisnik često i intenzivno koristi sistem, dok u suprotnom, ako su sesije kratke i neredovne, performanse sistema neće biti zadovoljavajuće.

Podaci koji se prikupljaju u procesu personalizacije mogu se svrstati u tri kategorije [Kob01]. To su:

- podaci o korisniku, koji se odnose na njegove lične karakteristike,
- podaci o korišćenju sistema, koji se odnose na interaktivno ponašanje korisnika i
- podaci o okruženju, koji se odnose na opis softvera i hardvera.

Posebnu podgrupu podataka o korišćenju sistema čine obrasci korišćenja, koji opisuju često ponavljane interakcije korisnika.

Detaljan prikaz navedenih kategorija podataka, s primerima sistema u kojima se koriste, dat je u tabeli 4.1. Iz nje vidimo da podaci koji su relevantni za personalizaciju variraju od sasvim opštih, kao što je, na primer, uzrast korisnika, do specifičnih, poput njegove trenutne lokacije. Takođe, za neke podatke je relevantna trenutna vrednost, dok se drugi odnose na sekvencu događaja i aktivnosti, posmatranu u dužem periodu vremena.

Često i detaljno prikupljanje podataka o korisniku sistema u kome je primenjena personalizacija pokreće pitanje zaštite kako njegove privatnosti, tako i ličnih podataka. Ovo pitanje je detaljno diskutovano u [Tel04], gde su sistematizovani rezultati opsežnih istraživanja (preko 30) sprovedenih nad korisnicima interneta, s naglaskom na servis elektronske kupovine. U nastavku ovoga odeljka, navećemo neke od njih:

- 70-83% ispitanika brine se o poverljivosti svojih ličnih podataka;

Tabela 4.1: Primeri podataka relevantnih za personalizaciju (preuzeto iz [Tel04]).

Podatak	Primer sistema
Demografski podaci	Personalizacija internet-sajtova zasnovana na profilima korisnika
Znanje korisnika	Personalizacija zasnovana na ekspertizi; tehnički opisi proizvoda; sistemi za učenje
Veštine i sposobnosti	Pomoć u korišćenju računarskih programa; pomoć korisnicima s posebnim potrebama
Interesovanja i preferencije	Sistemi za pružanje preporuka
Ciljevi i planovi	Personalizovana podrška korisnicima s definisanim ponašanjem
Selektivne akcije	Adaptacija zasnovana na izboru linka ili slike
Temporalno ponašanje	Adaptacija zasnovana na vremenu koje se utroši na interakciju sa sistemom
Ocenjivanje	Predlozi za kupovinu
Aktivnosti uz kupovinu	Predlaganje proizvoda sličnih kupljenom
Ostale aktivnosti	Adaptacija zasnovana na snimanju ili štampanju dokumenta
Učestanost korišćenja	Personalizacija menija i ikonica u računarskim programima
Sekvence akcija	Predlozi zasnovani na čestim sekvencama akcija, u prošlosti ili od drugih korisnika
Softversko okruženje	Adaptacija zasnovana na verziji pretraživača i operativnog sistema kod korisnika, instaliranim kodecima itd.
Hardversko okruženje	Adaptacija zasnovana na raspoloživom protoku konekcije, procesorskoj moći, memoriji, rezoluciji displeja itd.
Lokalne karakteristike	Adaptacija zasnovana na lokaciji korisnika (jezik, pozivni broj, valuta)

- 82% je barem jednom odbilo da ostavi svoje lične podatke na internet sajtu;
- 27% nikada ne bi ostavilo svoje lične podatke na internet sajtu;
- 24-34% je barem jednom ostavilo izmišljene ili lažne podatke na internet sajtu.

Poslednji rezultat je naročito zanimljiv, jer, uprkos onome što bi se moglo očekivati, ukazuje na opasnosti eksplicitnog modeliranja korisnika kroz popunjavanje upitnika prilikom registrovanja za korišćenje sistema.

Veoma zanimljivi rezultati odnose se na odnos pitanja privatnosti i učestanosti korišćenja sistema:

- Više od 60% ispitanika nikada nije koristilo elektronsku kupovinu, zbog straha od zloupotrebe svojih podataka.
- 20-37% korisnika elektronske kupovine bilo bi spremno da troši više novca kada bi bili uvereni da su njihovi podaci zaštićeni.

- 27% korisnika elektronske kupovine prestalo je s korišćenjem ovog servisa iz straha zbog zloupotrebe podataka.

Po pitanju opšteg stava prema personalizaciji, dobijeni su sledeći rezultati:

- 59% ispitanika se pozitivno izjašnjava o ideji personalizacije, a 37% negativno;
- 73% smatra da je dobro ako sistem „zapamti“ osnovne podatke o korisniku, kao što su, npr. ime i adresa;
- 50% smatra da bi sistem trebalo da pamti i dodatne podatke, kao što su, npr. omiljene robne marke, boje, opcije isporuke kupljene robe itd;
- 62% ispitanika ističe da im smeta kada sistem traži od njih da unesu podatke koje su već jednom uneli, kao npr. adresu isporuke.

Najveći broj korisnika spremjan je da sistemu daju podatak o svome imenu (88%), obrazovanju (88%), uzrastu (86%) i hobijima (83%). Vidimo da se tu radi o opštim podacima, koji imaju malu vrednost za proces formiranja modela korisnika i kasniju personalizaciju. Ukoliko se radi o iole „ličnjim“ podacima, korisnici izbegavaju da ih ostave sistemu (plata – 59%, broj kreditne kartice 13%), ili, kao što je ranije navedeno, umesto stvarnih ostavljaju lažne podatke.

Pitanje prikupljanja podataka o korisnicima i zaštite njihove privatnosti dobilo je i pravnu dimenziju. Direktivama evropskog parlamenta i Veća Evrope [EU95, EU02], propisano je da korisniku treba da bude omogućen:

- uvid u proces prikupljanja i obrade podataka,
- uvid u prikupljene podatke,
- mogućnost ispravljanja pogrešnih podataka,
- informacija o svrsi prikupljanja podataka.

Iz ovog pregleda vidimo da je pitanje zaštite privatnosti podataka od presudnog značaja za uspeh modeliranja korisnika i personalizacije. U literaturi još nema jedinstvenog stava o tome na koji način ono treba da se reši. Jedan mogući pristup bi se oslanjao na eksplicitno modeliranje korisnika, kojima bi se davale čvrste, po potrebi i zakonske garancije da se njihovi podaci neće zloupotrebiti. Korisnik bi pri tome imao punu kontrolu nad vrstom i količinom podataka koji se o njemu prikupljaju. Drugi pristup bi se zasnivao na unapređenju neinvazivnih tehnika implicitnog modeliranja korisnika. Ostaje da vidimo koji će od ovih – za sada oprečnih – pristupa prevagnuti.

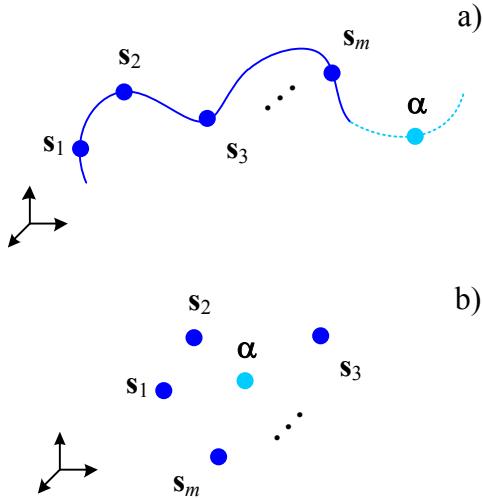
## 4.5 Modeliranje korisnika u vektorskim prostorima

U ovom odeljku ćemo pokazati na koji se način modeliranje korisnika može izvršiti u modelu vektorskog prostora.

Neka je dat vektorski prostor  $\Omega$ , dimenzije  $n$ , čiji su elementi telekomunikacioni servisi ili multimedijalni sadržaji i neka je poznat model korisnika,  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_n)$ . Ako

bismo u posmatranom prostoru definisali meru razdaljine  $\text{Dist}(\cdot)$ , ili sličnosti  $\text{Sim}(\cdot)$ , tada bismo mogli da svakom raspoloživom sadržaju iz njega pridružimo rastojanje ili sličnost u odnosu na aktuelni model korisnika. „Najbolji” sadržaj tada bi, očigledno, bio onaj s najmanjim rastojanjem ili s najvećom sličnošću; stoga se u kontekstu vektorskih prostora model korisnika može shvatiti i kao servis/sadržaj koji u najvećoj meri odgovara željama korisnika.

Neka je, dalje, poznato  $m$  servisa/sadržaja kojima je korisnik prethodno pristupao,  $\mathbf{s}_i = (s_{i,1}, s_{i,2}, \dots, s_{i,n})$ ,  $i = 1, \dots, m$ , pri čemu je  $\mathbf{s}_1$  najstariji, a  $\mathbf{s}_m$  najsukoriji. U prvi mah bismo mogli očekivati da će ovi sadržaji ležati na krivoj liniji koja bi odražavala evoluciju interesovanja posmatranog korisnika tokom vremena. Optimalni model korisnika tada bismo mogli odrediti ekstrapolacijom, što je ilustrovano na slici 4.3 a).



Slika 4.3: Određivanje optimalnog modela korisnika u vektorskome prostoru: a) ekstrapolacija, b) klasterovanje.

Teorija pronalaženja informacija, međutim, kaže da ovakvo očekivanje nije opravdano. Prema tzv. klasterskoj hipotezi, skup srodnih informacionih potreba (ili interesovanja) korisnika formira klaster, a ne krivu liniju [Rij79]; optimalni model korisnika u modelu vektorskog prostora stoga ćemo odrediti kao centroidu klastera sadržaja kojima je taj korisnik prethodno pristupao, što je simbolički prikazano na slici 4.3 b). Napomenimo da je važna pretpostavka pod kojom važi klasterska hipoteza to da su aktivnosti korisnika konzistentne, odnosno da nema naglih promena njegovih interesovanja.

Određivanje optimalnog modela korisnika sada možemo shvatiti kao optimizacioni problem minimizacije ukupnog rastojanja ili maksimizacije ukupne sličnosti, tj:

$$\boldsymbol{\alpha}_{opt} = \arg \min_{\boldsymbol{\alpha} \in \Omega} \sum_{i=1}^m \text{Dist} (\boldsymbol{\alpha}, \mathbf{s}_i), \quad (4.1)$$

ili

$$\boldsymbol{\alpha}_{opt} = \arg \max_{\boldsymbol{\alpha} \in \Omega} \sum_{i=1}^m \text{Sim} (\boldsymbol{\alpha}, \mathbf{s}_i). \quad (4.2)$$

**Primer 4.1.** Odredimo optimalni model korisnika za slučaj Euklidovog rastojanja.

*Rešenje.* Rastojanje sadržaja  $\mathbf{s}_j$  od modela korisnika,  $\boldsymbol{\alpha}$ , dano je izrazom

$$d_j = \sqrt{\sum_{i=1}^n (s_{j,i} - \alpha_i)^2}.$$

Potrebno je minimizirati funkciju

$$g(\boldsymbol{\alpha}) = \sum_{j=1}^m d_j = \sum_{j=1}^m \sqrt{\sum_{i=1}^n (s_{j,i} - \alpha_i)^2}.$$

Zbog monotonosti korene funkcije, originalni problem može se svesti na analitički jednostavniji problem minimiziranja funkcije

$$g_1(\boldsymbol{\alpha}) = \sum_{j=1}^m d_j = \sum_{j=1}^m \sum_{i=1}^n (s_{j,i} - \alpha_i)^2.$$

Odredimo parcijalne izvode:

$$\frac{\partial g_1(\boldsymbol{\alpha})}{\partial \alpha_i} = \sum_{j=1}^m (-2) (s_{j,i} - \alpha_i), \quad i = 1, \dots, n.$$

Njihovim izjednačavanjem s nulom, dobijamo koordinate optimalnog modela korisnika

$$\alpha_{i,opt} = \frac{1}{m} \sum_{j=1}^m s_{j,i}.$$

■

**Primer 4.2.** Odredimo optimalni model korisnika za slučaj kosinusne sličnosti nad normiranim vektorskim prostorom – ( $\forall \mathbf{x} \in \Omega$ )  $|\mathbf{x}| = 1$ .

*Rešenje.* Pod pretpostavkom normiranosti, kosinusna sličnost svodi se na skalarni proizvod:

$$\text{Sim}(\boldsymbol{\alpha}, \mathbf{s}_j) = \sum_{i=1}^n \alpha_i s_{j,i}.$$

Potrebno je maksimizirati funkciju

$$g(\boldsymbol{\alpha}) = \sum_{j=1}^m \text{Sim}(\boldsymbol{\alpha}, \mathbf{s}_j) = \sum_{j=1}^m \sum_{i=1}^n \alpha_i s_{j,i},$$

pod uslovom  $|\boldsymbol{\alpha}| = 1$ , koji je ekvivalentan sa

$$\sum_{i=1}^n \alpha_i^2 = 1.$$

Formirajmo ciljnu funkciju

$$\begin{aligned}\varphi(\boldsymbol{\alpha}, \beta) &= g(\boldsymbol{\alpha}) - \beta(|\boldsymbol{\alpha}|^2 - 1) = \\ &= \sum_{j=1}^m \sum_{i=1}^n \alpha_i s_{j,i} - \beta \left( \sum_{i=1}^n \alpha_i^2 - 1 \right),\end{aligned}$$

gde je  $\beta$  Lagrangeov multiplikator. Nakon nalaženja parcijalnih izvoda po  $\alpha_i$ ,  $i = 1, \dots, n$  i njihovog izjednjačavanja s nulom, dobijamo

$$\alpha_i = \frac{1}{2\beta} \sum_{j=1}^m s_{j,i}.$$

Vrednost  $\beta$  dobijamo nakon uvrštavanja u uslov normiranosti korisničkog modela:

$$2\beta = \sqrt{\sum_{k=1}^n \left( \sum_{j=1}^m s_{j,k} \right)^2}.$$

Konačno dobijamo

$$\alpha_{i,opt} = \frac{\sum_{j=1}^m s_{j,i}}{\sqrt{\sum_{k=1}^n \left( \sum_{j=1}^m s_{j,k} \right)^2}}.$$

■

## 4.6 Pitanja i zadaci

**4.1.** Diskutujte mogućnosti za eksplicitno i implicitno modeliranje korisnika, pa tabelarno navedite njihove dobre i loše strane.

**4.2.** Izvedite parametre optimalnog modela korisnika za slučaj Manhattan metrike.

**4.3.** Izvedite parametre optimalnog modela korisnika za slučaj zeta rastojanja.

**4.4.** Odredite red funkcije složenosti izračunavanja optimalnog normiranog modela korisnika za kosinusnu sličnost.

**4.5.** U tzv. algoritmu agregiranja, sistem ažurira dva zasebna podmodela – pozitivni ( $\boldsymbol{\alpha}^+$ ), koji opisuje šta se korisniku sviđa i negativni ( $\boldsymbol{\alpha}^-$ ), koji opisuje šta mu se ne sviđa. Na osnovu njih, formira se agregirani model

$$\boldsymbol{\alpha} = \mu \boldsymbol{\alpha}^+ - (1 - \mu) \boldsymbol{\alpha}^-,$$

gde je  $\mu$  koeficijent agregacije. Odredite njegovu vrednost tako da agregirani model bude ortogonalan s negativnim podmodelom. Pod kojim uslovom je to moguće?

# Poglavlje 5

## Personalizacija u praksi

### 5.1 Uvod

U ovome poglavlju detaljnije ćemo razmotriti aktuelne pristupe personalizaciji telekomunikacionih servisa.

Počećemo formalnom definicijom personalizacije. Personalizacija (telekomunikacionog) servisa podrazumeva da se korisniku  $U$  omogući da, u skladu sa svojim potrebama, napravi novi ili prilagodi postojeći servis  $S$ , tako da se nakon ovoga svi servisi koji proizilaze iz  $S$  ka  $U$  menjaju na isti način [Blo00]. Razradimo li ovu definiciju, možemo dalje reći da je cilj personalizacije prilagođavanje servisa potrebama i preferencijama ne samo pojedinačnog korisnika, već i grupe korisnika.

Koreni personalizacije u telekomunikacijama – i informacionim sistemima uopšte – leže u pojavi prvih adaptivnih korisničkih interfejsa, ličnih agenata i pomoćnika, te adaptivnog pronalaženja informacija [Yan06]. Funkcionalnosti u pogledu personalizacije razvijale su se tokom vremena, tako da je u današnjem potrošačkom društву, orijentisanom na korišćenje interneta i mobilnih telekomunikacija, personalizacija postala ključan faktor za prihvatanje telekomunikacionih servisa i aplikacija [Jør05].

Motivi za uvođenje personalizovanih aplikacija u osnovi se mogu svrstati u dve kategorije i to (1) olakšavanje rada i (2) prilagođavanje zahtevima društva [Blo00, Jør05]. U prvom slučaju, aplikacije se adaptiraju da bi se povećala efikasnost njihovog korišćenja, u smislu smanjenja vremena koje se troši na slične aktivnosti, koje se uz to i često ponavljaju. U ovu kategoriju spada i personalizacija koja korisnicima s invaliditetom (npr. s oštećenim slušom ili vidom) omogućava nesmetano korišćenje aplikacija. U drugom slučaju, aplikacije se adaptiraju s ciljem unapređivanja društvenog iskustva njihovih korisnika; tipičan primer pružaju korisnici mobilnih telefona, koji izborom melodija zvona i grafičkog prikaza na displeju izražavaju svoju posebnost.

Razmotrimo odnos personalizacije s modeliranjem korisnika i pronalaženjem informa-

cija. Podsetimo se slike 4.1. Personalizacija je usko povezana s modeliranjem korisnika. Proces personalizacije obuhvata prikupljanje podataka, njihovu obradu radi formiranja modela korisnika i korišćenje ovoga modela da bi se aplikacija prilagodila korisniku. Modeliranje korisnika samo za sebe, bez personalizacije koja bi za njim usledila, ne bi imalo svrhu.

Personalizacija je, s druge strane, povezana i s pronalaženjem informacija. Naime, personalizacija se može shvatiti i kao proces odlučivanja koja od datih alternativa ima najveću vrednost za pojedinca [Yan06]. Ovakvo tumačenje blisko je poimanju pronalaženja informacija, pa se, stoga, postavlja pitanje čemu uopšte personalizacija, ako se isti zadatak može obaviti i neposrednom primenom metoda i tehniku pronalaženja informacija [Val07]. Odgovor je da bi pronalaženje informacija, primenjeno za sebe, uvek davalо iste rezultate za isti upit, bez obzira na to o kom se korisniku radi. Udruženo s personalizacijom, pronalaženje informacija uzima u obzir karakteristike konkretnog korisnika i omogućava mu da nađe ili „skroji“ aplikaciju prema svojim željama.

U nastavku ovoga poglavlja, razmotreno je aktuelno stanje u trima grupama aplikacija koje su naročito zanimljive zbog potencijala za personalizaciju. To su:

1. sistemi za pružanje preporuka,
2. pretraživači interneta i
3. tzv. mobilne aplikacije.

## 5.2 Preporučivači sadržaja

Sistemi za pružanje preporuka (ili, kraće, preporučivači sadržaja) imaju za cilj procenu *ocena* koje bi njihovi korisnici dali objektima (npr. knjigama, filmovima, muzičkim datotekama, turističkim aranžmanima) koje nisu koristili. Sistem potom korisnicima može preporučiti objekte za koje su procenjene ocene najveće [Ado05].

Formalno, neka je  $U$  skup korisnika i  $I$  skup raspoloživih sadržaja. Definišimo preslikavanje

$$f : U \times I \mapsto R, \quad (5.1)$$

gde je  $R$  skup ocena, nenegativnih celih ili realnih brojeva iz ograničenog intervala. Za svakog korisnika  $u$ , cilj je naći objekat  $i'$  za koji je vrednost preslikavanja  $f$  maksimalna, tj.

$$(\forall u \in U) \quad i' = \arg \max_{i \in I} f(u, i). \quad (5.2)$$

Korisnici, po pravilu, ne ocenjuju sve raspoložive objekte, pa funkcija  $f$  nije definisana na celom prostoru  $U \times I$ , već na nekom njegovom podskupu. Kao što je to ilustrovano u tabeli 5.1, pridružena matrica ocena slabo je popunjena, pa se nedostajuće vrednosti određuju ekstrapolacijom raspoloživih.

Tabela 5.1: Primer matrice ocena.

	$i_1$	$i_2$	$i_3$	...
$u_1$	4	3	...	
$u_2$		5	1	...
$u_3$	2		3	...
:	:	:	:	..

U opštem slučaju, ova ekstrapolacija obuhvata dva koraka: u prvom se usvaja heuristika kojom se definiše funkcija  $f(u, i)$ , a u drugom se procenjuje njena vrednost za nedostajuće unose iz matrice ocena, tako da se optimizira zadati kriterijum performansi. Kada su poznate sve ocene – bilo tako što ih je dao sâm korisnik, ili je sistem procenio njihove vrednosti – korisniku se, prema (5.2) može ponuditi objekt s najvećom ocenom, ili, alternativno, nekoliko najbolje rangiranih, po nerastućem redosledu ocena, što podseća na princip rangiranja verovatnoće.

U zavisnosti od toga kako se izvode preporuke, tradicionalno se razlikuju *preporučivači zasnovani na opisu sadržaja* i *kolaborativni preporučivači*. Moguć je i kombinovan pristup, pa se tako dobijaju *hibridni preporučivači*. Odnedavno, u literaturi se sve više uviđa značaj konteksta pod kojim korisnici pristupaju sadržajima, pa se tako govori i o *preporučivačima s prepoznavanjem konteksta* ili o *trodimenzionalnim preporučivačima* [Ado11]. U nastavku ćemo razmotriti odlike svakog od ovih pristupa.

### 5.2.1 Sistemi zasnovani na opisu sadržaja

Preporučivači zasnovani na opisu sadržaja procenjuju nedostajuće ocene objekata na osnovu ocena koje je korisnik prethodno dao sličnim objektima. Ovakav pristup usko je povezan s pronalaženjem informacija i u osnovi odgovara onome što smo videli u odeljku 4.5: svaki objekt opisuje se svojim odlikama, npr. ključnim rečima, čije se „težine“ određuju npr. metodom TF-IDF (strana 30). Istovremeno, formira se model korisnika (str. 51) i potom se određuje sličnost neocenjenog objekta s njim (str. 38).

Personalizovani televizijski programski vodič TV3P [Yu04] s kojim smo se upoznali u primeru 3.2, ilustrativan je predstavnik ove klase preporučivača. U njemu je primenjeno predstavljanje televizijskih programa u modelu vektorskog prostora, uz korišćenje MPEG-7 *metadata* informacije. Model korisnika može se formirati trojako, eksplicitnim unošenjem parametara, implicitnom reakcijom ili eksplicitnom reakcijom.

Eksplicitno modeliranje je tehnički najjednostavnije i svodi se na to da korisnik prilikom registrovanja za korišćenje sistema izabere ključne reči koje ga zanimaju i zada njihov značaj. Izmena ovih parametara moguća je kasnije, u bilo kom trenutku korišćenja sistema.

Implicitna reakcija se zasniva na merenju vremena koje korisnik provede gledajući pre-

poručeni program. Ako neka odlika programa postoji u tekućem profilu korisnika, tada se njena težina ažurira po sledećim obrascima. Ako je  $w_i$  bila stara težina posmatrane odlike, njena nova težina je

$$w'_i = (1 - \alpha)w_i + \alpha\Delta w_i, \quad (5.3)$$

gde je  $\alpha \in [0, 1]$  koeficijent koji određuje brzinu učenja algoritma, a  $\Delta w_i$  promena težine, data kao

$$\Delta w_i = \beta f(i), \quad (5.4)$$

pri čemu je

$$\beta = \frac{T_r}{T_t} \in [0, 1] \quad (5.5)$$

količnik vremena koje korisnik provede gledajući preporučeni program i ukupnog trajanja tog programa, a

$$f(i) = \frac{I_{max} - i}{I_{max}}, \quad 1 \leq i \leq I_{max} \quad (5.6)$$

funcija koja opisuje značaj poretka odlika na preferencije korisnika. Podsetimo se da se odlike u opisu programa i modelu korisnika nalaze poredane po opadajućem značaju – značajnije odlike nalaze se na početku, a manje značajne na kraju modela; stoga  $f(i)$  opada s porastom rednog broja odlike unutar modela korisnika,  $i$ . Konačno,  $I_{max}$  je ukupan broj odlika u modelu korisnika.

Ukoliko posmatrana odlika preporučenog programa ne postoji u tekućem profilu korisnika, tada se njena težina računa po obrascu

$$w_i = \alpha\beta f(i), \quad (5.7)$$

pri čemu su značenja pojedinih veličina ista kao u prethodnom slučaju.

Ako je težina nove odlike veća od zadatog praga  $\lambda$ , smatra se da je ona značajna za korisnika i dodaje se njegovom modelu. U suprotnom, prepostavlja se da je nova odlika trivijalna i odbacuje se.

U slučaju eksplicitne reakcije, od interesa je je li korisnik prihvatio preporuku, ili ju je ignorisao, ili, možda, odbacio. Ako je korisnik eksplicitno prihvatio preporuku i prebacio se na novi kanal, tada je  $\Delta w_i = 2$ ; ako je ignorisao preporuku, usled čega se sistem sam prebacio na novi kanal, biće  $\Delta w_i = 1$ . Konačno, ako je korisnik odbio predlog, biće  $\Delta w_i = -2$ .

Primetimo da osnovna prednost ovakvog, tzv. *content-based* pristupa leži u jednostavnosti i očiglednosti; danas je sve značajnija i inherentna zaštita podataka o korisniku, jer se preporuke mogu izvoditi lokalno, na korisnikovom terminalu, bez slanja podataka o pristupanju sadržajima (*usage history*) nudiocu servisa. Mane su mu potreba za detaljnim opisom sadržaja, što smo diskutovali u odeljku 3.5, neraznovrsnost ponuđenih preporuka (prespecijalizacija), jer se one biraju tako da se podudaraju s procenjenim modelom korisnika i problem novog korisnika (hladni start) – kako ponudititi preporuku novom korisniku, koji nije generisao dovoljno interakciju da bi se formirao njegov model.

### 5.2.2 Sistemi zasnovani na kolaborativnom filtriranju

Kolaborativni preporučivači izvode preporuke na osnovu ocena koje su dali drugi korisnici, slični posmatranome. Stereotipi (str. 45) i zajednice (str. 46) tipični su primeri kolaborativnog filtriranja. U novije vreme, ovakvi preporučivači se koriste za ciljano reklamiranje na društvenim mrežama [Mer14].

Kolaborativne preporuke najčešće se izvode u dva koraka – u prvom se identifikuju korisnici koji su slični posmatranom, a u drugom se agregiraju njihove ocene da bi se procenila nedostajuća ocena objekta za posmatranog korisnika.

Neka je  $I_j$  skup objekata koje je ocenio korisnik  $u_j$ . Neka je, dalje, ocena koju je ovaj korisnik dao objektu  $i \in I_j$   $r_{j,i}$ , tako da je srednja ocena ovoga korisnika

$$\bar{r}_j = \frac{1}{|I_j|} \sum_{i \in I_j} r_{j,i}. \quad (5.8)$$

Uzajamna sličnost dvaju korisnika  $u_x$  i  $u_y$  u praksi se računa bilo kao Pearsonov koefficijent korelacije,

$$\text{sim}(u_x, u_y) = \frac{\sum_{i \in I_{xy}} (r_{x,i} - \bar{r}_x)(r_{y,i} - \bar{r}_y)}{\sqrt{\sum_{i \in I_{xy}} (r_{x,i} - \bar{r}_x)^2} \sqrt{\sum_{i \in I_{xy}} (r_{y,i} - \bar{r}_y)^2}}, \quad (5.9)$$

bilo kao kosinusna sličnost pridruženih im vektora,

$$\text{sim}(u_x, u_y) = \frac{\sum_{i \in I_{xy}} r_{x,i} r_{y,i}}{\sqrt{\sum_{i \in I_{xy}} r_{x,i}^2} \sqrt{\sum_{i \in I_{xy}} r_{y,i}^2}}, \quad (5.10)$$

pri čemu je  $I_{xy} = I_x \cap I_y$ .

Nakon što se formira klaster  $N$  korisnika najsličnijih korisniku  $u_j$ ,  $\hat{U}_j$ , njihove se ocene agregiraju po nekom od sledećih obrazaca:

$$r_{j,i} = \frac{1}{N} \sum_{u_k \in \hat{U}_j} r_{k,i}, \quad (5.11)$$

$$r_{j,i} = \varepsilon \sum_{u_k \in \hat{U}_j} \text{sim}(u_j, u_k) r_{k,i}, \quad (5.12)$$

$$r_{j,i} = \bar{r}_j + \varepsilon \sum_{u_k \in \hat{U}_j} \text{sim}(u_j, u_k) (r_{k,i} - \bar{r}_k), \quad (5.13)$$

$$\text{gde je } \varepsilon = \frac{1}{\sum_{u_k \in \hat{U}_j} \text{sim}(u_j, u_k)}.$$

**Primer 5.1.** U tabeli su date ocene koje su korisnici A–E dali objektima  $i_1$ – $i_5$ . Odredimo uzajamne sličnosti ovih korisnika pomoću Pearsonovog koeficijenta korelacije.

Tabela 5.2: Uz primer 5.1.

	A	B	C	D	E
$i_1$	9	10	7	10	9
$i_2$	7	6	2	1	1
$i_3$	5	1	5	5	4
$i_4$	3	5	3	2	2
$i_5$	1	3	5	7	6

*Rešenje.* Neposrednom primenom izraza (5.9), dobijamo rezultate iz sledeće tabele.

Tabela 5.3: Uzajamne sličnosti korisnika.

	A	B	C	D	E
A	1	0,699	0,243	0,215	0,246
B	0,699	1	0,265	0,341	0,413
C	0,243	0,265	1	<b>0,977</b>	0,669
D	0,215	0,341	<b>0,977</b>	1	<b>0,996</b>
E	0,246	0,413	0,669	<b>0,996</b>	1

Primetimo da je matrica sličnosti simetrična u odnosu na glavnu dijagonalu; vrednosti s nje u praksi se često izostavljaju, jer odgovaraju trivijalnom slučaju maksimalne sličnosti korisnika sa samom sobom. Pri interpretaciji rezultata ipak treba biti oprezan, jer ova simetrija ne važi i za rangiranje korisnika po sličnostima. Kao što se vidi iz primera, korisniku C najsličniji je D, ali korisniku D nije najsličniji C, već E. ■

Osnovna prednost kolaborativnog pristupa je to što u njemu nema potrebe za opisom sadržaja; doista, objekti se ovde tretiraju kao apstrakcije opisane svojim jedinstvenim identifikatorom u kolekciji. Kolaborativno filtriranje stoga je pogodno za primenu u situacijama u kojima nije moguća automatska analiza sadržaja [Cha14]. Problem hladnog starta može se donekle rešiti svrstavanjem korisnika u zajednicu, na osnovu njegovih socio-demografskih podataka. Novost u ovom pristupu je problem dodavanja novog objekta, koji korisnici nisu imali priliku da ocene i koji se, stoga, nikome ne može ni preporučiti. Problem predstavljaju i atipični korisnici, tzv. crne ovce, za koje se ne mogu naći parnjaci, kao i velike dimenzije matrice ocena koju treba obraditi.

### 5.2.3 Hibridni sistemi

Hibridni sistemi kombinuju dva prethodno razmotrena pristupa, opis sadržaja i kolaborativno filtriranje. Neke od mogućnosti su:

- ocene se izvode kao ponderisana suma ocena dobijenih primenom opisa sadržaja i kolaborativnog filtriranja,
- sličnost između korisnika se računa na osnovu njihovih profila,
- na grupu modela korisnika, koji su predstavljeni vektorima, primenjuje se neka od metoda smanjenja dimenzionalnosti prostora da bi se dobio kolaborativni profil [The15] itd.

Poslednji metod zanimljiv je kako sa stanovišta izvođenja kvalitetnijih preporuka, tako i sa stanovišta smanjenja složenosti algoritma. Tako su, na primer, Isobe i dr. pokazali da se televizijski programi mogu opisati trima odlikama koje odgovaraju opuštajućem, informativnom i emotivnom sadržaju [Iso05]. Sledeći ovu ideju, na uzorku studenata Elektrotehničkog fakulteta odredili smo koeficijente transformacione matrice za različite žanrove televizijskih programa, koji su prikazani u tabeli 5.4 [Krs12].

Tabela 5.4: Koeficijenti transformacione matrice.

	Relaxing	Informative	Emotional
Action	24	47	29
Adventure	25	46	29
Animation	33	47	20
Children's	25	47	28
Comedy	24	48	28
Crime	24	49	27
Documentary	24	28	48
Drama	22	50	28
Fantasy	22	51	27
Film noir	22	52	26
Fun	85	5	10
Horror	23	50	27
Lifestyle	72	19	9
Musical	25	48	27
Music show	52	0	48
Mystery	25	47	28
News/info	0	95	5
Romance	25	46	29
Sci-Fi	24	47	29
Sport	21	19	60
Talk show	40	1	59
Thriller	24	46	30
War	23	48	29
Western	23	51	26

U poslednje vreme, sve je zanimljiviji tzv. jedinstveni pristup, koji inherentno objedinjuje opis sadržaja i kolaborativno filtriranje. Jedan ovakav način za procenu ocena opisan je sledećim obrascima [Ado05]:

$$r_{u,i} = x_{u,i}\mu + z_u\gamma_i + w_i\lambda_u + e_{u,i}, \quad (5.14)$$

$$e_{u,i} \sim \mathcal{N}(0, \sigma^2), \quad (5.15)$$

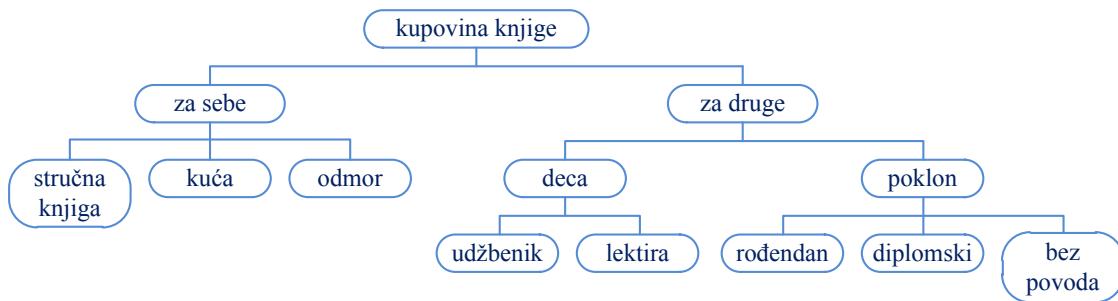
$$\lambda_u \sim \mathcal{N}(0, \Lambda), \quad (5.16)$$

$$\gamma_i \sim \mathcal{N}(0, \Gamma), \quad (5.17)$$

gde indeks  $u$  označava korisnika, a indeks  $i$  objekt iz kolekcije;  $e_{u,i}$ ,  $\lambda_u$  i  $\gamma_i$  su slučajne promenljive kojima se redom opisuju uticaji šuma, te neosmotrivoih uzroka heterogenosti korisnika i objekata. Dalje,  $x_{u,i}$  je matrica s opisom afiniteta korisnika prema objektima,  $z_u$  je vektor odlika korisnika, a  $w_i$  vektor odlika objekata. Nepoznati parametri modela su  $\mu$ ,  $\sigma^2$ ,  $\Lambda$  i  $\Gamma$  i njihove se vrednosti određuju na osnovu raspoloživih – poznatih – ocena.

#### 5.2.4 Sistemi zasnovani na prepoznavanju konteksta

Najšire shvaćen, kontekst predstavlja uslove i okolnosti koji utiču na neku aktivnost. Ako je, na primer, aktivnost kupovina knjige, izbor naslova će se razlikovati u zavisnosti od toga kupujemo li knjigu za sebe, ili za nekoga drugog; u potonjem slučaju, knjiga može biti namenjena deci, pa se verovatno radi ili o školskom udžbeniku, ili o lektiri primerenoj uzrastu i interesovanjima deteta, ili se, pak, može raditi o poklonu, pri čemu sad konkretan povod utiče na ishod kupovine. Ovo je ilustrovano na slici 5.1, u vidu grafovske strukture. Idući od korena grafa, koji predstavlja aktivnost, prema njegovim listovima, dobijamo sve vredniju informaciju o kontekstu pod kojim se ta aktivnost obavlja.



Slika 5.1: Hijerarhijska struktura konteksta.

Sistemi zasnovani na prepoznavanju konteksta [Ado11] predstavljaju novi pravac u razvoju preporučivača. Oni izvode preporuke tako što posmatraju i kontekst pod kojim korisnici pristupaju raspoloživim objektima. Imajući u vidu izraz (5.1), koji opisuje funkcionisanje „tradicionalnih“ preporučivača, za kontekstualne možemo pisati

$$f : U \times I \times C \mapsto R, \quad (5.18)$$

gde je C skup koji sadrži kontekstualne informacije. Zbog toga što imaju trojake ulazne podatke – korisnike, objekte i kontekst, za kontekstualne preporučivače se kaže i da su trodimenzionalni.

Ključan element za funkcionisanje kontekstualnih preporučivača je akvizicija informacije o kontekstu. Kao i kod formiranja modela korisnika (str. 45), za to postoji više načina. Kontekst se, pre svega, može spoznati eksplicitnim modeliranjem, npr. popunjavanjem upitnika. Nasuprot tome, implicitna varijanta podrazumeva akviziciju podataka sa senzora korisničkog terminala, npr. vreme interakcije se dobija očitavanjem sistemskog sata. Konačno, kontekst se može dobiti i zaključivanjem, tj. obradom podataka nižeg reda; tako se, na primer, iz sekvence televizijskih programa kojima je pristupano na zajedničkom televizijskom prijemniku može pretpostaviti o kom se članu domaćinstva radi. Sve što smo prethodno kazali za modeliranje korisnika, direktno se može primeniti i ovde.

Korišćenjem konteksta mogu se ostvariti značajno bolje performanse preporučivača bez povećanja neprijatnosti po korisnika i uz neznatno usložnjavanje sistema [Krs12].

### 5.3 Pretraživači interneta

Iako možda na prvi pogled tako ne izgleda, pretraživanje interneta izuzetno je zahtevno zbog ogromne količine raspoloživih dokumenata. Iskusan korisnik moći će da formuliše svoj upit pomoću odgovarajućeg rečnika i tako će dobiti linkove ka traženim objektima; neiskusan korisnik će, s druge strane, često imati problem ili da formuliše svoj zahtev za informacijom, ili da se snađe u mnoštvu rezultata pretraživanja. Zvuči paradoksalno, ali sasvim je moguće da problem za korisnika ne bude izostanak rezultata pretraživanja, već njihov prevelik broj.

Naredna specifičnost interneta je postojanje hiperlinkova između dokumenata, koji uzrokuju pojavu vezanih informacija. Dok s jedne strane to može povoljno uticati na zadovoljavanje korisnikove potrebe za informacijama, s druge može i otežati samo pronalaženje inicijalne informacije. Konačno, dokumenti koji su raspoloživi na internetu izuzetno su heterogene strukture. Pored različitih medija (teksta, slika, animacija, audio i video klipova) i formata, internet sajtovi često sadrže i delove (npr. reklame) koji se ne odnose direktno na njihov osnovni sadržaj.

U literaturi je opisano pet obrazaca traženja informacija na internetu [Meg08]. To su:

1. skeniranje (oblast se pokriva plitko, u širinu),
2. pretraživanje (sledi se konzistentna putanja dok se ne bude dostigao cilj),
3. traganje (traženje dok se ne bude dostigao eksplicitan cilj),
4. istraživanje (zadata informacija se ispituje u dubinu) i
5. lutanje (besciljno i nestrukturirano traženje).

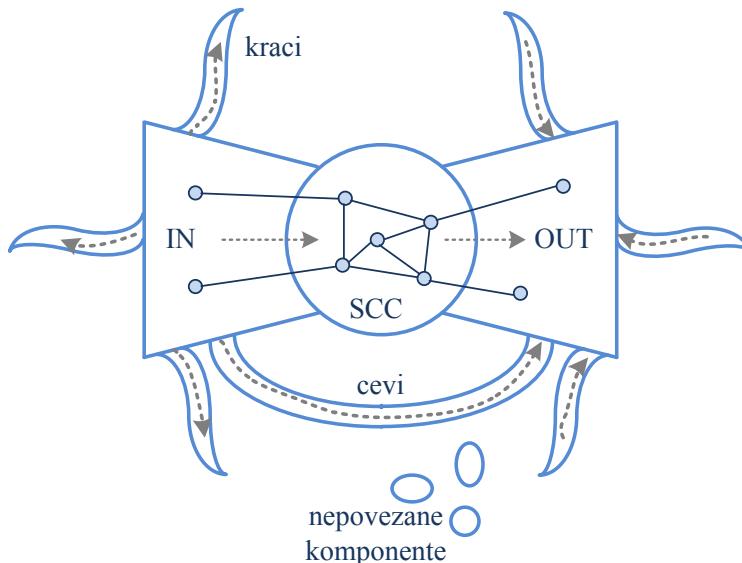
Tri važna rezultata do kojih se došlo istraživanjem postupka pretraživanja interneta

su:

- Tipičan korisnik interneta nastoji da maksimizira količinu informacija koje dobije tokom sesije [Pat07].
- Tipičan upit nema više od tri reči. Pri njegovom formulisanju, retko se koriste logički operatori [Pat07].
- Korisnici po pravilu pogledaju samo nekoliko prvih linkova koje je pretraživač pronašao, a ostale zanemaruju [Jan98].

Poslednji zaključak govori da korisnici interneta prihvataju suboptimalno rešenje i zbog toga propuštaju globalno optimalno; ova je pojava opšta karakteristika pretraživanja velikih skupova dokumenata.

Za razumevanje rada pretraživača interneta, potrebno je da se upoznamo s njegovom topologijom. U sada već klasičnom radu [Bro00], pokazano je da struktura interneta, posmatrana u vidu orijentisanog grafa čiji su čvorovi internet stranice, a grane linkovi koji ih povezuju, podseća na leptir-mašnu, što je ilustrovano na slici 5.2.



Slika 5.2: Grafovska struktura interneta.

„Čvor” mašne odgovara dobro povezanim stranicama u jezgru interneta (SCC). U ovoj oblasti, korisnici interneta lako se kreću s jednog sajta na drugi. Na levoj strani su izvorišne stranice (IN) s kojih korisnici mogu doći do jezgra, ali obrnut put nije moguć; primer su nove stranice, koje još nisu privukle pažnju internet zajednice. Na desnoj strani su odredišne stranice (OUT), do kojih se može doći iz jezgra, ali put u obrnutom smeru nije moguć; tipični primjeri bile bi komercijalne i korporacijske stranice.

S bilo koje stranice iz dela IN može se, preko jezgra (SCC) doći do dela OUT. Postoje, međutim, i stranice do kojih se može doći samo iz delova oblasti IN, kao i one koje vode samo do delova oblasti OUT; one su na slici označene kao kraci. Spajanjem dvaju kraka dobijaju se cevi. Konačno, postoje i regioni nepovezanih komponenti, koje

tipično sadrže opskurne i privatne sajtove.

U tabeli 5.5, upoređene su veličine pojedinih segmenata interneta prema podacima dobijenim 1999. i 2012. godine.

Tabela 5.5: Veličine segmenata topologije interneta (u %).

Komponenta	1999. [Bro00]	2012. [Meu14]
SCC	27,74	51,28
IN	21,29	31,96
OUT	21,21	6,05
kraci	21,52	4,61
cevi	—	0,26
nepovezano	8,24	5,84

Iako konkretni algoritmi nisu javno dostupni, izvesno je da komercijalni pretraživači određuju relevantnost internet stranice za upit kombinovanim vrednovanjem, pomoću tehnika pronalaženja informacija, posmatranjem konteksta u kome se postavlja upit, ali takođe i na osnovu „kvaliteta“ stranice koja je kandidat za uključivanje na listu pronađenih. Potonji parametar je jasan u svetu slike 5.2, jer je za očekivati da će stranice ka kojima vodi više linkova korisnicima biti zanimljivije. U pretraživaču Google primjenjen je *PageRank* algoritam [Bri98], u kome se polazi od modela korisnika koji se kreće s jedne stranice na drugu tako što na slučajan način bira neki od dostupnih linkova; ovakav korisnik u literaturi se naziva *random surfer*. „Kvalitet“ stranice, ili njen *rang* odgovara verovatnoći da će *random surfer* doći do te stranice.

Za zadati graf internet stranica koji sadrži  $N$  čvorova, definiše se matrica linkova  $\mathbf{A} = [a_{i,j}]$ ,  $i, j = 1, \dots, N$ , takva da je

$$a_{i,j} = \begin{cases} \frac{1}{n_j}, & \text{ako postoji grana od čvora } j \text{ ka čvoru } i, \\ 0 & \text{u suprotnom} \end{cases}, \quad (5.19)$$

pri čemu je  $n_j$  broj grana koje izlaze iz čvora  $j$ . Primetimo da je  $a_{i,i} = 0$ . Verovatnoća da će *random surfer* posetiti stranicu  $i$  tada je

$$x_i = a_{i,1}x_1 + a_{i,2}x_2 + \dots + a_{i,N}x_N, \quad i = 1, 2, \dots, N, \quad (5.20)$$

što se u matričnom obliku može zapisati kao

$$\mathbf{x} = \mathbf{Ax}. \quad (5.21)$$

Sada je trenutak da se prisjetimo linearne algebre. Sopstveni vektori  $\mathbf{x}$  i sopstvene vrednosti  $\lambda$  matrice  $\mathbf{A}$  zadovoljavaju jednačinu  $\mathbf{Ax} = \lambda\mathbf{x}$ , pa se traženi vektor verovatnoća,  $\mathbf{x}$ , može shvatiti kao sopstveni vektor matrice linkova za jediničnu sopstvenu vrednost.

Jednačina (5.21) omogućava iterativno izračunavanje vektora verovatnoća, odnosno rangova stranica. Doista, budemo li obe njene strane množili sleva matricom linkova, imaćemo

$$\mathbf{x} = \mathbf{A}\mathbf{x} = \mathbf{A}^2\mathbf{x} = \mathbf{A}^3\mathbf{x} = \dots, \quad (5.22)$$

što znači da bismo mogli pretpostaviti inicijalne vrednosti elemenata vektora  $\mathbf{x}$ , npr.  $1/N$  i potom primeniti iterativni postupak

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k, \quad k \in \mathbb{N}_0. \quad (5.23)$$

Može se pokazati da opisani postupak konvergira ka jedinstvenom nenegativnom rešenju ukoliko je  $\mathbf{A}$  stohastična, iderucibilna i aperiodična matrica [Bry06].

**Primer 5.2.** Odredimo rangove čvorova u grafu sa slike 5.3.



Slika 5.3: Uz primer 5.2.

*Rešenje.* Za ovaj jednostavni graf, očigledno je da rang čvora 1 treba da bude nula, a čvora 2 jedan. Proverimo to neposrednim izračunavanjem: matrica linkova je

$$\mathbf{A} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Prepostavimo inicijalni vektor rangova  $\mathbf{x}_0 = [0,5 \ 0,5]^T$ . Tada će biti  $\mathbf{x}_1 = [0 \ 0,5]^T$  i  $\mathbf{x}_2 = \mathbf{x}_3 = \dots = [0 \ 0]^T$ , pa postupak konvergira ka pogrešnom rešenju. ■

Da bi se povećala klasa matrica za koju postupak konvergira ka tačnom rešenju, umesto matrice  $\mathbf{A}$  posmatra se

$$\mathbf{M} = d\mathbf{A} + \frac{1-d}{N}\mathbf{1}, \quad (5.24)$$

gde je  $\mathbf{1}$  matrica dimenzija  $N \times N$  čiji su elementi jedinice i  $0 \leq d \leq 1$ ; vrednost ove konstante koju koristi pretraživač Google prema indirektnim saznanjima iznosi 0,85. Ovako definisana, matrica  $\mathbf{M}$  naziva se Google matricom. Matrica verovatnoća ili rangova internet stranica,  $\mathbf{x}$ , sada se računa kao rešenje jednačine

$$\mathbf{x} = \mathbf{Mx}, \quad (5.25)$$

ili, ekvivalentno,

$$\mathbf{x} = d\mathbf{Ax} + \frac{1-d}{N}\mathbf{1}_{N \times 1}, \quad (5.26)$$

gde je  $\mathbf{1}_{N \times 1}$  vektor-kolona s  $N$  jedinica. U literaturi se sreće i alternativna definicija, kao

$$\mathbf{x} = d\mathbf{Ax} + (1-d)\mathbf{1}_{N \times 1}. \quad (5.27)$$

Skalarni zapis izraza (5.26) bio bi

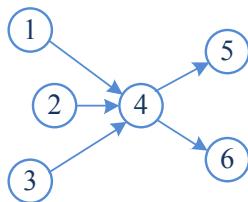
$$x_i = \frac{1-d}{N} + d \sum_{j \in M_i} \frac{x_j}{n_j}, \quad (5.28)$$

a (5.27)

$$x_i = (1 - d) + d \sum_{j \in M_i} \frac{x_j}{n_j}, \quad (5.29)$$

gde je  $M_i$  skup čvorova iz kojih postoje grane ka čvoru  $i$  i  $i, j \in \{1, 2, \dots, N\}$ .

**Primer 5.3.** Odredimo rangove čvorova u grafu sa slike 5.4.



Slika 5.4: Uz primer 5.3.

*Rešenje.* Inspekcijom grafa, zaključujemo da je  $x_1 = x_2 = x_3 = \alpha$  i  $x_5 = x_6 = \beta$ . Primenom obrasca (5.28), dobijamo  $\alpha = \frac{0.15}{6} = 0,025$ ,  $x_4 = 0,08875$  i  $\beta \approx 0,06272$ . Najveći je rang čvora 4, a najmanji čvorova 1, 2 i 3. ■

## 5.4 Aplikacije za mobilne terminalne

Mobilni terminali su od čisto radiotelefonskih postali napredni IT uređaji, koji se koriste za telefoniju u užem smislu, razmenjivanje tekstualnih, slikovnih i video poruka, razmenjivanje elektronske pošte, pristup internetu, obradu dokumenata, fotografisanje i snimanje video-klipova, slušanje radija i muzike, gledanje video sadržaja itd. Zbog popularnosti i jednostavnijeg korišćenja u odnosu na klasične računare, mobilni uređaji stvoriti su ogroman prostor za personalizaciju aplikacija i sadržaja koje se nude njihovim korisnicima.

Pored opštih odlika koje smo već razmotrili, personalizacija aplikacija koje se izvršavaju na mobilnim terminalima ima i dve specifičnosti. Pre svega, ovde je potrebno prilagoditi servis još uvek ograničenim hardverskim i softverskim resursima mobilnog terminala, kao što su procesorska moć, kapacitet memorije, veličina displeja, broj i veličina tastera, kapacitet baterije, operativni sistem itd. [Hua03]. Druga odlika proizilazi iz same mobilnosti korisnika, kojima je potrebno obezbediti neometano obavljanje ostalih aktivnosti, npr. kretanje ulicom ili vožnju javnim prevozom, dok se aplikacija izvršava.

Istraživanja su pokazala da se korisnici prema mobilnim terminalama odnose na krajnje lični način [Ich05]. Doista, retki su korisnici koji nisu „personalizovali“ svoj mobilni

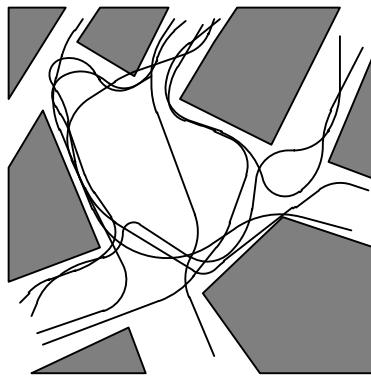
telefon, kroz melodije zvona, pozadine displeja, posebne futrole itd. Nećemo trošiti prostor na objašnjavanje lukrativnosti tržišta mobilnih telekomunikacija; umesto toga, samo ćemo izneti podatak da je u sada već dalekoj 2006. godini ostvaren neto prihod od 4 milijarde USD na preuzimanju melodija zvona za mobilne telefone [Alc07].

Naročito su zanimljive aplikacije koje se adaptiraju prema procenjenoj fizičkoj lokaciji korisnika; one čine osnovu takozvanih lokacijskih servisa (engl. *Location-Based Services*, LBS) [Sch04], čiji su tipični primeri turistički vodiči, vodiči za kupovinu i izliske, ciljane reklame itd. Primetimo da se precizan podatak o lokaciji može dobiti kako pomoću satelitskih navigacionih sistema (GPS), tako i pomoću mobilne mreže, nekom od metoda pozicioniranja. U praksi je, naročito u urbanim sredinama, sasvim dovoljno poznavati i okvirnu lokaciju, koja se može jednostavno dobiti očitavanjem identifikacije servisne bazne stanice [Che12].

U nastavku odeljka, predstavićemo osnove jednog LBS sistema [Bau05]. Položaj korisnika se određuje GPS prijemnicima, infracrvenim senzorima ili WLAN senzorima. Svako merenje  $p_i$  predstavlja uređenu četvorku  $(x_i, y_i, z_i, t_i)$ , čija su prva tri člana prostorne koordinate korisnika, a četvrti trenutak merenja.

Profil kretanja je par  $(modp(S), modm(S))$ , u kome je  $S = (p_1, p_2, \dots, p_n)$  sekvenca merenja. Prvi član profila kretanja predstavlja apstrakciju lokacija iz  $S$ , a drugi apstrakciju samog kretanja. Cilj modeliranja kretanja je pružanje preporuka korisniku koje su za njega potencijalno interesantne. To je moguće bilo kolaborativnim preporučivanjem, upoređivanjem profila kretanja s profilima ostalih korisnika, ili predikcijom na osnovu strukturalnih odlika korisnikovog kretanja.

Kolaborativne preporuke izvode se tako što se snimaju sekvene kretanja unutar jedne oblasti (slika 5.5) i potom određuju klasteri srodnih kretanja. Oni se potom zamenjuju medioidama, krivim linijama koje imaju najmanje prosečno rastojanje od svih elemenata klastera.



Slika 5.5: Sekvence kretanja u urbanoj sredini.

Rastojanje sekvenci kretanja  $S_1$  i  $S_2$  dato je izrazom

$$\text{dist}_S(S_1, S_2) = w_e \min (\text{dist}_e(S_1, S_2), \text{dist}_e(S_2, S_1)) + w_0 \text{disorder}(S_1, S_2). \quad (5.30)$$

Ovde su  $w_e$  i  $w_0$  težinski koeficijenti,  $\text{dist}_e$  je tzv. edit rastojanje sekvenci, a disorder broj pozicija u  $S_1$  koje su u obrnutom redosledu u odnosu na  $S_2$ . Edit rastojanje preuzeto je iz poređenja stringova i odgovara minimalnom broju operacija umetanja, brisanja i zamene koje treba izvršiti da bi prvi argument postao jednak drugom. Pri izvršavanju svake od ovih operacija, plaća se određena cena.

Izdvajanjem medioida i upoređivanjem profila kretanja različitih korisnika unutar jedne prostorne oblasti moguće je uočiti pravilnosti u kretanju i formulisati preporuke. Na primer, navodi se da je na uzorku od 145 korisnica koje su posetile samoposlužu i prodavnicu obuće, 27% ušlo i u restoran.

Predikcija kretanja korisnika vrši se analizom sekvence  $S$ , tako što se identifikuju sledeće pojave:

- postojanje petlji, njihovo trajanje, brzina kretanja, učestanost na predenom kilometru,
- zadržavanje na pojedinim lokacijama, prosečno trajanje, tip posećenih mesta,
- skretanja s prvobitne putanje, pri čemu se pravi razlika između namernih i obaveznih (npr. zbog skretanja ulice).

Opisani koncept bio je implementiran u nemačkom gradu Saarbrückenu, gde su se putem WLAN korisnicima dostavljale personalizovane informacije o raspoloživim sadržajima u gradu, kao i ciljane reklamne poruke.

## 5.5 Pitanja i zadaci

**5.1.** Uporedite sisteme za pružanje preporuka zasnovane na opisu sadržaja s kolaborativnim – tabelarno navedite njihove dobre i loše strane. Za svaki navedeni nedostatak predložite način na koji bi se mogao otkloniti.

**5.2.** Diskutujte mogućnosti za eksplicitno i implicitno modeliranje gledalaca televizije, pa tabelarno navedite njihove dobre i loše strane.

**5.3.** Model (profil) korisnika u sistemu TV3P [Yu04] je

```
<Term Weight=1.25>romance</Term>
<Term Weight=1.16>animal</Term>
<Term Weight=1.10>love</Term>
<Term Weight=1.00>hero</Term>
<Term Weight=0.84>marriage</Term>
<Term Weight=0.78>sport</Term>
<Term Weight=0.72>church</Term>
<Term Weight=0.60>gun</Term>.
```

Odredite (korak po korak) sličnost ovoga profila i filma čiji je *metadata* opis

```
<Title>Gone with the Wind</Title>
<Genre>Romance</Genre>
<Actor>Clark Gable, Vivien Leigh</Actor>
<Keyword>Love</Keyword>
<Keyword>Marriage</Keyword>
<Keyword>War</Keyword>
<Duration unit=minute>180</Duration>.
```

**5.4.** Kombinovani eksplicitno-implicitni model korisnika u sistemu TV3P [Yu04] formira se po obrascu

$$w'_i = (1 - \alpha)w_i + \alpha(W\_I \cdot \Delta w_i\_I + W\_E \cdot \Delta w_i\_E),$$

pri čemu se  $\Delta w_i\_I$  računa iz implicitne reakcije, a  $\Delta w_i\_E$  iz eksplicitne.  $W\_I$  i  $W\_E$  su težinski faktori kojima se vrednuju doprinosi implicitne i eksplicitne reakcije, respektivno i za koje važi  $W\_I + W\_E = 1$ .

Neka je  $\alpha = 0,25$ ,  $W\_I = 0,4$  i  $W\_E = 0,6$ . Neka se korisniku iz prethodnog zadatka pored filma *Gone with the Wind* nudi i *Animal World*, čiji je opis

```
<Title>Animal World</Title>
<Genre>Animal</Genre>
<Keyword>Nature</Keyword>.
```

Posmatrajte sledeći scenario dešavanja: Korisnik na ovu preporuku odgovara tako što odbija *Animal World* (pritiskom tastera *Remove*), a prihvata *Gone with the Wind*. Odredite sadržaj profila korisnika posle 160 minuta gledanja ovoga filma.

**5.5.** Korisnik je ocenio objekte  $i_1, i_2, i_3$  i  $i_4$  redom ocenama 5, 3, 1 i 2. Neka se u sistem uvode novi objekti  $i_5$  i  $i_6$ , koje korisnik još nije ocenio i čije su respektivne sličnosti s objektima  $i_1, \dots, i_4$  0,8, 0,9, 0,7, 0,3 (za  $i_5$ ) i 0,7, 0,4, 0,9, 0,9 (za  $i_6$ ). Procenite vrednosti novih objekata za ovoga korisnika.

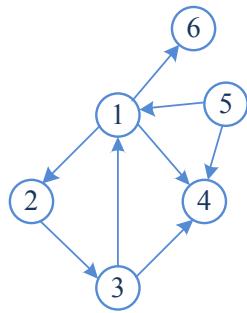
**5.6.** Broj dokumenata koji su raspoloživi na internetu neprekidno raste; stoga se performanse pretraživača interneta ne mogu izraziti preko preciznosti i sposobnosti pronalaženja, onako kako smo ih definisali u odeljku 2.8. Umesto njih, ovde se definišu njihovi pandani  $P_{@20}$  i  $C_{@20}$ , na sledeći način [Pat05]:  $P_{@20}$  je deo prvih 20 pronađenih linkova koji su relevantni za korisnika, a  $C_{@20}$  se definiše izrazom

$$C_{@20} = \min\left(\frac{P_{@20}E}{20}, 1\right),$$

gde je sa  $E$  označen procenjeni broj relevantnih dokumenata u bazi pretraživača.

Po uzoru na primer 2.5, konstruišite upit i rezultate pretrage, te za njih odredite vrednosti  $P_{@20}$  i  $C_{@20}$ .

**5.7.** Odredite rangove čvorova u grafu sa slike.



**5.8.** Proverite rezultat primera [5.3](#) nalaženjem sopstvenog vektora Google matrice.



# Dodatak: Personalizovani programski vodiči

## Uvod

Na kraju ove knjige, osvrnućemo se na projektovanje personalizovanih programskih vodiča; ova oblast predmet je užeg stručnog interesovanja autora, a poslužiće nam kao ilustrativna studija slučaja modernih personalizovanih aplikacija.

Većini gledalaca, ogroman broj programa osnovna je odlika digitalne, pogotovo kablov-ske televizije. Istovremeno, to izobilje sadržaja sobom nosi i frustraciju za korisnike servisa, jer im je često teško da nađu program koji bi ih interesovao. Kao i kod pretraživanja interneta (odeljak 5.3), problem nije u nedostatku željenih informacija, već u mnoštvu irelevantnih, koje maskiraju željene i otežavaju njihovo pronalaženje. Sistematski uvid u raspoložive programe – redom jedan po jedan – zahteva značajno vreme i trud, te stoga nije prihvatljiv, baš kao što nije prihvatljiv ni uvid u rasporede programa, bilo u štampanoj, ili elektronskoj varijanti. Mnogi mogu potvrditi iz ličnog iskustva da na kraju, od velikog broja televizijskih programa, po pravilu redovno gledaju svega nekoliko – u proseku, dugoročno ne više od devet [Bje08], dok dok ostale zanemaruju.

Moguće rešenje ovoga problema bili bi personalizovani programski vodiči, tj. softverski agenti koji gledaocima televizije preporučuju sadržaje od interesa. Vredi istaći da njihova primena nije ograničena samo na klasičnu televiziju; veoma perspektivno polje otvara se s porastom lične video produkcije i pojmom internet platformi za distribuciju ovakvih sadržaja.

## Zahtevi i ograničenja

Projektanti personalizovanih programskih vodiča na samom se početku susreću s prilično striktnim zahtevima u pogledu kvaliteta finalnog proizvoda i ograničenjima koja imaju na putu do njega.

Televizija je primer telekomunikacionog servisa s duboko izgrađenom navikom korišćenja.

Personalizovani programski vodiči tu su da pomognu svojim korisnicima da ovo korišćenje bude još priyatnije i ni u kom slučaju ne bi smeli da ih u tome ometaju. Ovo se direktno odražava na vrstu i tok dijaloga sistema s korisnikom. Specijalizovani hardver, poput računarskih tastatura, namenskih daljinskih upravljača i ekrana osetljivih na dodir, nije dostupan i pitanje je hoće li uopšte biti u doglednoj budućnosti; isti komentar važi i za komandu putem pokreta ili glasa. Ono što preostaje je standardni daljinski upravljač, eventualno s par dodatnih tastera; ovakvo rešenje doista je dobro poznato korisniku i ne bi od njega tražilo napor u cilju učenja, privikavanja i kasnijeg korišćenja.

Zahtevanje eksplisitne reakcije korisnika putem davanja numeričkih ocena emisiji koja se prikazuje predstavljačko bi nepotrebno ometanje u korišćenju servisa; s druge strane, izražavanje stava preko dvaju tastera na daljinskom upravljaču („Like” i „Dislike”) izgleda kao dobro polazno rešenje. U literaturi su opisani i predlozi za praćenje biometrijskih parametara (pulsa, frekvencije disanja, pokreta očiju, položaja tela i sl), no oni zahtevaju intenzivno korišćenje namenskih – i skupih – senzora, čije nošenje na telu ili odeći većini korisnika verovatno ne bi bilo prijatno. Takođe, izvođenje zaključka na osnovu toga koliko je dugo korisnik gledao neki program diskutabilne je vrednosti; lako se može desiti da korisnik neposredno pre odlaska na posao pogleda samo najvažnije vesti, iz čega bi ovakav sistem pogrešno zaključio da njegovog korisnika vesti ne zanimaju.

Klasična televizija je širokodifuzni servis, što znači da se informacije prenose od centra ka periferiji, tj. korisnicima; ne postoji povratni kanal, kojim bi se informacija prenela u obrnutom smeru. Dok novi televizijski servisi, poput IPTV ili DVB-H raspolažu povratnim kanalom, veliko je pitanje može li se na njega uvek računati. Čak i kada tehnološke mogućnosti to dozvoljavaju, opravdano je pretpostaviti da će biti korisnika koji će želeti da ga isključe, npr. radi zaštite svoje privatnosti, smanjenja izloženosti elektromagnetskom zračenju, ili produženja veka baterije, ukoliko servisu pristupaju s mobilnih terminala. Odsustvo povratnog kanala nameće potrebu korišćenja algoritama zasnovanih na modeliranju sadržaja, jer kolaborativni nisu primenljivi.

Konačno, treba voditi računa o hardverskim resursima koji su dostupni za izvršavanje aplikacije vodiča, naročito u slučaju mobilnih terminala (odeljak 5.4). U principu, poželjna su rešenja što manje algoritamske složenosti.

## Primer 1: vodič kao preporučivač sadržaja

Televizijski programi predstavljaju se normiranim vektorima odllika, na osnovu pridruženih *metadata* ili ETT (*Extended Text Tables*) podataka [Bje10].

Označimo s  $\Omega$  skup raspoloživih programa. Kao mera uzajamne sličnosti programa  $\mathbf{p}_i, \mathbf{p}_j \in \Omega$  koristi se kosinusna sličnost, koja se zbog uslova normiranosti svodi na oblik

$$\text{Sim}(\mathbf{p}_i, \mathbf{p}_j) = \frac{\mathbf{p}_i \cdot \mathbf{p}_j}{|\mathbf{p}_i| |\mathbf{p}_j|} = \sum_{u=1}^n p_{i,u} p_{j,u}.$$

Za program  $\mathbf{p}_j$  kaže se da je u  $\delta$ -okolini programa  $\mathbf{p}_i$  ukoliko njihova sličnost nije manja od  $\delta$ :

$$K_\delta(\mathbf{p}_i) = \{\mathbf{p}_j \mid \mathbf{p}_j \in \Omega \wedge \text{Sim}(\mathbf{p}_i, \mathbf{p}_j) \geq \delta\}.$$

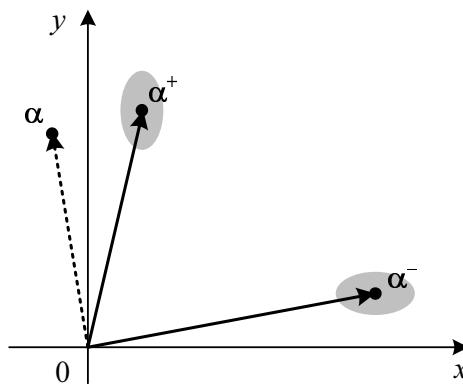
Korisniku se svaki put nude tri preporuke, a on iskazuje reakciju na sadržaj koji gleda kroz tzv. implicitno pozitivnu shemu [Bje11a], tako što u slučaju nezadovoljstva pritiska taster „Dislike” na daljinskom upravljaču, dok u suprotnom, ukoliko mu se program sviđa, ili je ambivalentan, ne preduzima nikakvu reakciju. Pozitivno ocenjeni programi čuvaju se u skupu  $P^+$ , a negativni u  $P^-$ ; na osnovu njih, formiraju se najpre pozitivni ( $\alpha^+$ ) i negativni ( $\alpha^-$ ) podmodel korisnika, respektivno, pa se potom određuje agregirani model

$$\alpha = \mu\alpha^+ - (1 - \mu)\alpha^-,$$

gde je

$$\mu = \begin{cases} \frac{1}{1 + \alpha^+ \cdot \alpha^-}, & \alpha^+ \cdot \alpha^- \leq \zeta, \\ \mu_0, & \text{u suprotnom} \end{cases}$$

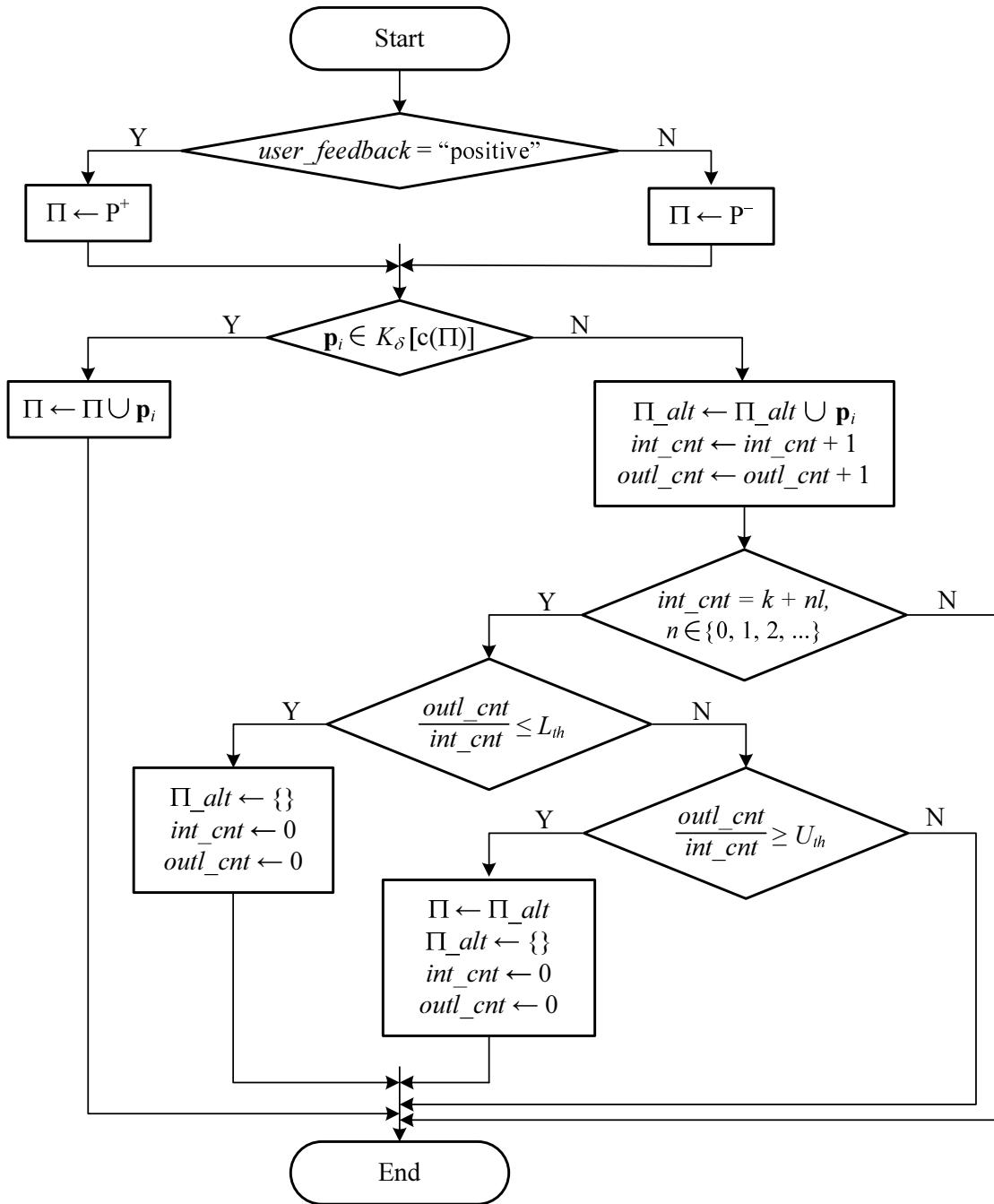
koeficijent agregacije i  $\zeta$  pozitivna konstanta. Predloženi postupak agregiranja motivisan je transformacijom upita iz pronalaženja informacija, (3.27); onda kada su afiniteti korisnika jasno definisani, razlika pozitivnog i negativnog podmodela preći će zadati prag i agregirani se model može formirati tako da bude ortogonalan na negativni podmodel. U suprotnom, kada je razlika između dvaju podmodела mala, aggregaciona konstanta uzima predefinisanu vrednost, jer ortogonalna agregacija ovde ne bi bila opravdana – agregirani bi model doista bio ortogonalan s negativnim podmodelom, ali bi se značajno udaljio i od pozitivnog.



Geometrijsko tumačenje agregiranog modela korisnika.

Za očekivati je da će se korisnikove preferencije menjati tokom vremena. Za adekvatno praćenje ovih promena može se primeniti sledeći adaptivni algoritam [Bje07, Bje11b]. Sve dok su interakcije konzistentne s važećim podmodelima, radi se o finim promenama ili tzv. driftu interesovanja. U ovome slučaju, dovoljno je redovno ažurirati model, tako da obuhvati  $m$  poslednjih interakcija, koje su raspoređene u skupove  $P^+$  i  $P^-$ . Nekonzistentna interakcija ukazuje na značajniju promenu interesovanja; takva interakcija čuva se u tzv. alternativnom podmodelu, pa se nove preporuke izvode po principu 2 (stari model) + 1 (alternativni). Kroz mehanizam brojača ukupnih ( $int\_cnt$ ) i nekonzistentnih reakcija ( $outl\_cnt$ ), kao i donjeg ( $L_{th}$ ) i gornjeg praga ( $U_{th}$ ), utvrđuje se je li ova promena interesovanja kratkotrajna – npr. ako korisnik koga prethodno nije

naročito interesovala politika počinje da prati vesti samo da bi saznao rezultate izbora, ili, pak, suštinska – korisnik, po preporuci, počinje da prati program koji ga ranije nije interesovao. Odluka se donosi nakon  $k + nl$  interakcija, gde su  $k$  i  $l$  predefinisane konstante, a  $n$  brojač, po principu histereze – ukoliko je u ovome intervalu relativna frekvencija nekonzistentnih interakcija manja od donjeg praga, zaključuje se da je promena interesovanja bila privremena i alternativni model se odbacuje. Ukoliko je, pak, relativna frekvencija nekonzistentnih interakcija veća od gornjeg praga, smatra se da je promena interesovanja trajna, te alternativni model postaje glavni. Konačno, ako nijedan od ova dva uslova nije ispunjen, nastavljaju se paralelno održavati oba modela i nova odluka se odlaže za još  $l$  interakcija.



Algoritam za praćenje promene korisnikovih interesovanja.

Sistem je testiran za sledeće vrednosti parametara:  $k = 4$ ,  $l = 2$ ,  $L_{th} = 0,25$ ,  $U_{th} = 0,75$ ,  $\zeta = 0,25$  i  $\mu = 1/(1 + \zeta) = 0,8$ . Za  $m = 20$ , ostvareno je 91% uspeha u prihvatanju preporuke, pri čemu se eksplicitna reakcija korisnika tražila u 24% slučajeva.

## Primer 2: vodič kao binarni klasifikator

Projektovanje personalizovanog programske vodiča može se posmatrati i kao problem binarne klasifikacije, u kome je svakom ulaznom objektu – televizijskom programu – potrebno dodeliti jednu od kategorija „like” ili „dislike”, koja odgovara procenjenom stavu korisnika prema tome objektu. Ovakav preporučivač može se efikasno implementirati preko neuronske mreže s jednim skrivenim slojem, čiji su ulazi opisi programa u vektorskom prostoru redukovane dimenzije (tabela 5.4), te kontekstualni podaci o vremenu kada je korisnik gledao taj program (radni dan, vikend, praznik; jutro, podne, popodne, veče, noć) [Krs12].

Ukoliko vodič dobro radi, svojim će korisnicima nuditi sadržaje koji bi trebalo da im se svide; zbog toga će broj negativnih interakcija biti daleko manji od pozitivnih, pa su potrebni efikasni algoritmi za treniranje klasifikatora kako bi uspešno mogao da prepozna neželjene sadržaje [Krs15]. U konkretnom slučaju, optimalni izlazni koeficijenti mreže trenirane na  $P$  interakcija, s  $K$  čvorova u skrivenom sloju iznose

$$\hat{\beta} = \begin{cases} \mathbf{H}^T \left( \frac{1}{R} \mathbf{I} + \mathbf{B} \mathbf{H} \mathbf{H}^T \right)^{-1} \mathbf{B} \mathbf{Z}, & P \leq K \\ \left( \frac{1}{R} \mathbf{I} + \mathbf{H}^T \mathbf{B} \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{B} \mathbf{Z}, & P > K \end{cases},$$

gde je  $\mathbf{H}$  matrica izlaza skrivenog sloja,  $\mathbf{Z}$  matrica željenih izlaza,  $R$  pozitivna konstanta i  $\mathbf{B}$  dijagonalna matrica, čiji element  $b_{ii}$ ,  $i = 1, \dots, P$  za pozitivnu interakciju iznosi  $0,618/\text{mes}\{\mathcal{P}^+\}$ , a za negativnu  $1/\text{mes}\{\mathcal{P}^-\}$ . Primetimo da vrednost 0,618 odgovara zlatnom preseku.

U ekstremnom slučaju kada nema negativnih interakcija, neophodno je koristiti tzv. unarne algoritme klasifikacije [Krs16].



# Literatura

- [Ado05] Gediminas Adomavicius and Alexander Tuzhilin: “Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions”, *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734–749, June 2005.
- [Ado11] Gediminas Adomavicius and Alexander Tuzhilin: “Context-aware recommender systems”, in *Recommender Systems Handbook*, Springer US, pp. 217–253, 2011.
- [Agg01] C. C. Aggarwal, A. Hinneburg, and D. A. Keim: “On the surprising behavior of distance metrics in high dimensional spaces”, in *ICDT ’01: Proceedings of the 8th International Conference on Database Theory*, London, UK: Springer-Verlag, pp. 420–434, 2001.
- [Alc07] Izvor: SRG NET, Alcatel Lucent Communication Day, CD ROM, Belgrade, November 2007.
- [Ard04] Liliana Ardissono, Cristina Gena, Pietro Torasso, Fabio Bellifemine, Angelo Difino and Barbara Negro: “User modeling and recommendation techniques for personalized electronic program guides”, in L. Ardissono et al. (eds.), *Personalized Digital Television*, pp. 3–26, Kluwer Academic Publishers, 2004.
- [Bae99] Ricardo Baeza-Yates and Berthier Ribeiro-Neto: “Modern Information Retrieval”, Addison-Wesley, 1999.
- [Bal08] E. Balmashnova, L. M. J. Florack: “Novel similarity measures for differential invariant descriptors for generic object retrieval”, *Journal of Mathematical Imaging and Vision*, volume 31, pp. 121–132, June 2008.
- [Bau05] Mathias Bauer and Matthieu Deru: “Motion-based adaptation of information services for mobile users”, *Lecture Notes in Computer Science*, vol. 3538, pp. 271–276, 2005.
- [Bec03] Jörg Becker and Dominik Kuropka: “Topic-based vector space model”, in Witold Abramowicz, Gary Klein (eds.): *Proceedings of BIS 2003*, Colorado Springs, USA, 2003.
- [Bje07] Milan Bjelica and Zoran Petrović: “A novel service retrieval scheme”, *IEEE Communications Letters*, vol. 11, no. 7, pp. 637–639, July 2007.

- [Bje08] Milan Bjelica: “Experiment with user modeling for communication service retrieval”, *IEEE Communications Letters*, vol. 12, no. 10, pp. 797–799, August 2008.
- [Bje10] Milan Bjelica: “Towards TV recommender system: Experiments with user modeling”, *IEEE Transactions on Consumer Electronics*, vol. 56, no. 3, pp. 1763–1769, August 2010.
- [Bje11a] Milan Bjelica: “Unobtrusive relevance feedback for personalized TV program guides”, *IEEE Transactions on Consumer Electronics*, vol. 57, no. 2, pp. 658–663, May 2011.
- [Bje11b] Milan Bjelica and Ana Perić: “Adaptive feedback schemes for personalized content retrieval”, *IEEE Transactions on Consumer Electronics*, vol. 57, no. 3, pp. 1251–1257, August 2011.
- [Blo00] Jan Blom: “Personalization — a taxonomy”, *CHI – Conference on Human Factors in Computing Systems*, the Hague, the Netherlands, 2000.
- [Bou06] Mokrane Bouzeghoub and Dimitre Kostadinov: “Data personalization: A taxonomy of user profiles knowledge and a profile management tool”, Report #2006/90, Laboratoire PriSM, Université de Versailles, 2006.
- [Bri98] Sergey Brin, Lawrence Page: “The anatomy of a large-scale hypertextual Web search engine”, *Computer Networks and ISDN Systems*, vol. 30, no. 1–7, pp. 107–117, April 1998.
- [Bro00] Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, Janet Wiener: “Graph structure in the Web”, *Computer Networks*, vol. 33, pp. 309–320, 2000.
- [Bry06] Kurt Bryan and Tanya Leise: “The \$25,000,000,000 eigenvector: The linear algebra behind Google”, *SIAM Review*, vol. 48, no. 3, pp. 569–581, March 2006.
- [Bus45] Vannevar Bush: “As we may think”, *Atlantic Monthly*, no. 176, pp. 101–108, July 1945.
- [Car87] John M. Carroll and Mary Beth Rosson: “Paradox of the active user”, in J. M. Carroll (ed.), *Interfacing Thought: Cognitive Aspects of Human-Computer Interaction*, pp. 80–111, MIT Press, Cambridge, MA, 1987.
- [Cha14] Ghislaine Chartron and Gérald Kembellec: “General introduction to recommender systems”, in Gérald Kembellec, Ghislaine Chartron and Imad Saleh (eds.): *Recommender Systems*, ISTE Wiley, 2014.
- [Che95] Hsinchun Chen: “Machine learning for information retrieval: neural networks, symbolic learning, and genetic algorithms”, *Journal of the American Society for Information Science*, vol. 46, no. 3, pp. 194–216, 1995.

- [Che05] Sen-Ching S. Cheung and Avideh Zakhori: “Fast Similarity Search and Clustering of Video Sequences on the World-Wide-Web”, *IEEE Transactions on Multimedia*, vol. 7, no. 3, pp. 524–538, June 2005.
- [Che12] Ruizhi Chen: “Ubiquitous Positioning and Mobile Location-Based Services in Smart Phones”, IGI Global, 2012.
- [Cre95] Fabio Crestani and Cornelis J. van Rijsbergen: “Information retrieval by logical imaging”, *Journal of Documentation*, vol. 51, pp. 3–17, 1995.
- [Cro94] Valerie Cross: “Fuzzy information retrieval”, *Journal of Intelligent Information Systems*, vol. 3, no. 1, pp. 29–56, February 1994.
- [Dee90] Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman: “Indexing by latent semantic analysis”, *Journal of the American Society for Information Science*, vol. 41, no. 6, pp. 391–407, 1990.
- [Dou03] Nikolaos D. Doulamis, Anastasios D. Doulamis, and Theodora A. Varvarigou: “Adaptive algorithms for interactive multimedia”, *IEEE Multimedia*, vol. 10, pp. 38–47, October-December 2003.
- [Dub04] David Dubin: “The most influential paper Gerard Salton never wrote”, *Library Trends*, vol. 52, no. 4, pp. 748–764, 2004.
- [EU02] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications), *Official Journal of the European Communities*, 31.07.2002.
- [EU95] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, *Official Journal of the European Communities*, 23.11.1995.
- [God05] Daniela Godoy and Analía Amandi: “User profiling for Web page filtering”, *IEEE Internet Computing*, pp. 56–64, July-August 2005.
- [Gót05] Júlia Góth, Adrienn Skrop: “Varying retrieval categoricity using hyperbolic geometry”, *Information Retrieval*, vol. 8, pp. 265–283, 2005.
- [Hua03] Tse-Hua Lan and Ahmed H. Tewfik: “A resource management strategy in wireless multimedia communications — Total power saving in mobile terminals with a guaranteed QoS”, *IEEE Transactions on Multimedia*, vol. 5, no. 2, pp. 267–281, June 2003.
- [Ich05] F. Ichikawa, J. Chipchase, and R. Grignani: “Where’s the phone? A study of mobile phone location in public spaces”, *Proceedings of the IEEE Mobility Conference*, Guangzhou, China, 2005.

- [Iso05] Tadashi Isobe, Masao Fujiwara, Hiroyuki Kaneta, Toshiya Morita, and Noriyoshi Uratani: “Development of a TV reception navigation system personalized with viewing habits”, *IEEE Transactions on Consumer Electronics*, vol. 51, no. 2, pp. 665–674, May 2005.
- [Ito05] Y. Ito and S. Tosaka: “Quantitative assessment of user-level QoS and its mapping”, *IEEE Transactions on Multimedia*, vol. 7, pp. 572–585, June 2005.
- [ITU01] ITU-T Recommendation G.1000: “Communications Quality of Service: A Framework and Definitions”, ITU-T 11/2001.
- [Jan98] B. J. Jansen, A. Spink, J. Bateman, T. Saracevic: “Real life information retrieval: A study of user queries on the web”, *SIGIR Forum*, vol. 32, no. 1, pp. 5–17, 1998.
- [Jør05] Ivar Jørstad, Do van Thanh and Schahram Dustdar: “The personalization of mobile services”, *IEEE International Conference on Wireless and Mobile Computing, Networking and Communications*, Vol. 4, pp. 59–65, August 2005.
- [Kay04] Judy Kay and Andrew Lum: “Building user models from observations of user accessing multimedia learning objects , in A. Nürnberger and M. Detyniecki (eds.), *Adaptive Multimedia Retrieval: First International Workshop, Hamburg, Germany: Revised Selected and Invited Papers, Lecture Notes in Computer Science 3094*, pp. 36–57, 2004.
- [Kel03] D. Kelly, J. Teevan, “Implicit feedback for inferring user preference: A bibliography”, *SIGIR Forum*, vol. 32, no. 2, 2003.
- [Kem02] C. Kemp and K. Ramamohanarao, “Long-term learning for web search engines”, in T. Elomaa et al. (eds.), *Lecture Notes in Artificial Intelligence 2431*, pp. 263–274, Springer-Verlag Berlin Heidelberg, 2002.
- [Kob01] A. Kobsa, “Generic user modeling systems”, *User Modeling and User-Adapted Interaction*, vol. 11, pp. 49–63, 2001.
- [Kos07] Markus Koskela, Alan F. Smeaton, and Jorma Laaksonen: “Measuring concept similarities in multimedia ontologies: Analysis and evaluations”, *IEEE Transactions on Multimedia*, vol. 9, no. 5, pp. 912–922, August 2007.
- [Krs12] Marko Krstić and Milan Bjelica: “Context-aware personalized program guide based on neural network”, *IEEE Transactions on Consumer Electronics*, vol. 58, no. 4, pp. 1301–1306, Nov. 2012.
- [Krs15] Marko Krstić and Milan Bjelica: “Impact of class imbalance on personalized program guide performance”, *IEEE Transactions on Consumer Electronics*, vol. 61, no. 1, pp. 90–95, Feb. 2015.
- [Krs16] Marko Krstić and Milan Bjelica: “Personalized program guide based on one-class classifier”, *IEEE Transactions on Consumer Electronics*, vol. 62, no. 2, pp. 175–181, May 2016.

- [Kur04] Dominik Kuropka: „Modelle zur Repräsentation natürlichsprachlicher Dokumente: Ontologie-basiertes Information-Filtering und -Retrieval mit relationalen Datenbanken”, Logos Verlag, Berlin, 2004.
- [Lee97] Dik L. Lee, Huei Chuang, and Kent Seamons: “Document ranking and the vector-space model”, *IEEE Software*, March/April 1997, pp. 67–75.
- [lib07] Internet sajt [www.libraries.gr](http://www.libraries.gr) – sadržaj preuzet septembra 2007.
- [Mar60] M. E. Maron and J. L. Kuhns: “On relevance, probabilistic indexing and information retrieval”, *Journal of the ACM*, vol. 7, no. 3, pp. 216–244, July 1960.
- [Meg08] George Meghabghab, Abraham Kandel: “Search Engines, Link Analysis and User’s Web Behaviour”, Springer-Verlag, Berlin Heidelberg, 2008.
- [Mer14] Maria Mercanti-Guérin: “Recommender systems and social networks: What are the implications for digital marketing?”, in Gérald Kembellec, Ghislaine Chartron and Imad Saleh (eds.): *Recommender Systems*, ISTE Wiley, 2014.
- [Meu14] Robert Meusel, Sebastiano Vigna, Oliver Lehmburg, Christian Bizer: “Graph Structure in the Web — Revisited, or A Trick of the Heavy Tail”, *Proceedings of the 23rd International Conference on World Wide Web*, pp. 427–432, 2014.
- [Mun06] P. Muneesawang, L. Guan: “Multimedia Database Retrieval: A Human-Centered Approach”, Springer Science, Boston/Dordrecht/London, 2006.
- [Nic93] W. K. Nicholson: “Elementary Linear Algebra With Applications”, PWS Publishing, Boston, 1993.
- [Orw95] J. Orwant: “Heterogeneous learning in the Doppelgänger user modeling system”, *User Modeling and User-Adapted Interaction*, vol. 4, no. 2, pp. 107–130, 1995.
- [Pap01] Christos Papageorgiou: “Machine learning in user modeling”, in G. Paliouras, V. Karakaletsis, and C. D. Spyropoulos (eds.), *Advanced Course in Artificial Intelligence ’99, Lecture Notes in Artificial Intelligence 2049*, pp. 286–294, 2001.
- [Pat05] Sunada Patro, Vishv Malhotra: “Characteristics of the Boolean Web search query: Estimating success from characteristics”, *Proceedings of the First International Conference on the Web Information Systems and Technologies*, pp. 339–344, Miami, USA, May 2005.
- [Pat07] Sunada Patro, Vishv Malhotra, and David Johnson: “An algorithm to use feedback on viewed documents to improve Web query: Enabling naïve searchers to search the Web smartly”, in J. Filipe, J. Cordeiro, and V. Pedrosa (eds.), *WEBIST 2005/2006, Lecture Notes in Business Information Processing*, vol. 1, pp. 177–189, 2007.

- [Pet06] Zoran Petrović i Milan Bjelica: „Novi pristupi personalizaciji telekomunikacionih servisa”, *Zbornik radova 24. simpozijuma o novim tehnologijama u poštanskom i telekomunikacionom saobraćaju*, str. 309–318, Saobraćajni fakultet, Beograd, 2006.
- [Poh95] Wolfgang Pohl, Alfred Kobsa and Oliver Kutter: “User model acquisition heuristics based on dialogue acts”, in *International Workshop on the Design of Cooperative Systems*, pp. 471–486, Antibes-Juan-Les-Pins, 1995.
- [Poh96] W. Pohl, “Learning about the user — User modeling and machine learning”, in V. Moustakis and J. Herrmann (eds.), *Proc. ICML ’96 Workshop “Machine Learning Meets Human-Computer Interaction”*, pp. 29–40, 1996.
- [Poh99] Wolfgang Pohl, Ingo Schwab, and Ivan Koychev: “Learning about the user: A general approach and its application”, in *IJCAI ’99 Workshop on Learning about Users*, Swedish Institute of Computer Science, 1999.
- [Pon98] Jay M. Ponte and W. Bruce Croft: “A language modeling approach to information retrieval”, *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 275–281, Melbourne, Australia, 1998.
- [Rag86] V. V. Raghavan and S. K. M. Wong: “A critical analysis of vector space model for information retrieval”, *Journal of the American Society for Information Science*, vol. 37, pp. 279–287, September 1986.
- [Rij79] C. J. Van Rijsbergen: “Information Retrieval”, Butterworths, London, 1979.
- [Ram03] Juan Ramos: “Using TF-IDF to Determine Word Relevance in Document Queries”, *The First Instructional Conference on Machine Learning, iCML 2003*, Piscataway, USA, 2003.
- [Rob76] S. E. Robertson, K. Spärck Jones: “Relevance weighting of search terms”, *Journal of the American Society for Information Science*, vol. 27, no. 3, pp. 129–146, May-June 1976.
- [Sal63] G. Salton: “Associative document retrieval techniques using bibliographic information”, *Journal of the Association for Computing Machinery*, vol. 10, no. 4, pp. 440–457, 1963. c.f. [Dub04]
- [Sal75] G. Salton, A. Wong, and C. S. Yang: “A vector space model for automatic indexing”, *Communications of the ACM*, vol. 18, no. 11, pp. 613–620, 1975.
- [Sal83a] Gerard Salton, Edward A. Fox, and Harry Wu: “Extended Boolean information retrieval”, *Communications of the ACM*, vol. 26, no. 11, pp. 1022–1036, November 1983.
- [Sal83b] G. Salton and M. J. McGill: “Introduction to Modern Information Retrieval”, McGraw-Hill, New York, USA, 1983. c.f. [Dub04]

- [Sal89] G. Salton: “Automatic Text Processing: The transformation, analysis, and retrieval of information by computer”, Addison-Wesley, Massachusetts, USA, 1989.
- [Sch04] Jochen Schiller, Agnès Voisard: “Location-Based Services”, Morgan Kaufmann Publishers, 2004.
- [Sin01] Amit Singhal: “Modern information retrieval: A brief overview”, *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, 2001.
- [Smy04] Barry Smyth and Paul Cotter: “Case-studies on the evolution of the personalized electronic program guide” in L. Ardissono et al. (eds.), *Personalized Digital Television*, pp. 53–71, Kluwer Academic Publishers, 2004.
- [Spä72] Karen Spärck Jones: “A statistical interpretation of term specificity and its application in retrieval”, *Journal of Documentation*, vol. 28, no. 1, pp. 11–21, 1972.
- [Tam05] L. Tamine, M. Boughanem, N. Zemirli: “Learning the user’s interests using the search history”, *NIPS 2005: Workshop on Machine Learning for Implicit Feedback and User Modeling*, Whistler, Canada, 2005.
- [Tel04] Maximilian Teltzrow, Alfred Kobsa: “Impacts of user privacy preferences on personalized systems”, in C.-M. Karat, J. Blom and J. Karat (eds.), *Designing Personalized User Experiences for eCommerce*, Kluwer Academic Publishers, Dordrecht, Netherlands, 2004.
- [The15] Sergios Theodoridis: “Machine Learning: A Bayesian and Optimization Perspective”, Elsevier, 2015.
- [Tse04] Belle L. Tseng, Ching-Yung Lin, and John R. Smith: “Using MPEG-7 and MPEG-21 for personalizing video”, *IEEE Multimedia*, pp. 42–52, January–March 2004.
- [Tur89] Howard Turtle and W. Bruce Croft: “Inference networks for documents retrieval”, *Proceedings of the 13th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1–24, Brussels, Belgium, 1989.
- [Val07] David Vallet, Pablo Castells, Miriam Fernández, Phivos Mylonas, and Yannis Avrithis: “Personalized content retrieval in context using ontological knowledge”, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 3, pp. 336–346, March 2007.
- [Vol07] Timo Volkmer, James A. Thom, and Seyed M. M. Tahaghoghi: “Modeling human judgement of digital imagery for multimedia retrieval”, *IEEE Transactions on Multimedia*, vol. 9, no. 5, pp. 967–974, August 2007.
- [Wah86] Wolfgang Wahlster and Alfred Kobsa: “Dialogue-based user models”, *Proceedings of the IEEE*, vol. 74, no. 7, pp. 948–960, July 1986.

- [Web01] Geoffrey I. Webb, Michael J. Pazzani and Daniel Billsus: “Machine learning for user modeling”, *User Modeling and User-Adapted Interaction*, vol. 11, pp. 19–29, 2001.
- [Won95] S. K. M. Wong, Wojciech Ziarko, and Patrick C. N. Wong: “Generalized vector spaces model in information retrieval”, *Proceedings of the 8th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 18–25, Montreal, Canada, 1995.
- [Yan06] Yuping Yang: “Provisioning of personalized pervasive services: Daidalos personalization functions”, *IEEE International Symposium on Pervasive Computing and Applications*, pp. 110–115, 2006.
- [Yu04] Zhiwen Yu and Xingshe Zhou: “TV3P: An adaptive assistant for personalized TV”, *IEEE Transactions on Consumer Electronics*, vol. 50, no. 1, pp. 393–399, February 2004.
- [Zim04] John Zimmerman, Kaushal Kurapati, Anna L. Buczak, Dave Schaffer, Sriniwas Gutta and Jacquelyn Martino: “TV personalization system: Design of a TV show recommender engine and interface”, in L. Ardissono et al. (eds.), *Personalized Digital Television*, pp. 27–51, Kluwer Academic Publishers, The Netherlands, 2004.
- [Zob98] Justin Zobel, Alistar Moffat: “Exploring the similarity space”, *ACM SIGIR Forum*, vol. 32, no. 1, pp. 18–34, 1998.