Sistemi preporuke

* Sistem preporuke zasnovan na kolaborativnom filtriranju, konkretno na pristupu opisanom u radu “Course Recommender System based on Graduating Attributes”**[[1]](#footnote-1)**. Pristup nije nužno vezan za edukativni domen, uz manje modifikacije može se primeniti na druge domene.
* Content-based sistem preporuke za preporuku knjiga na osnovnu sličnosti tema identifikovanih u opisu sadržaja knjiga.
  + Za prikupljanje podataka o opisima knjiga, koristiti npr. OCLC API[[2]](#footnote-2) ili neki drugi API za pristup podacima o knjigama[[3]](#footnote-3)
  + Za identifikaciju tema u opisima, koristiti neku od topic modelling metoda (npr. LDA)
  + Za računanje sličnosti knjiga na osnovu identifikovanih tema, može se koristiti *Hellinger distance* koju Blei and Lafferty (2009)[[4]](#footnote-4) preporučuju za računanje sličnosti dokumenata predstavljenih preko distribucije tema.
  + Umesto opisa knjiga, mogu biti i opisi nekih drugih artikala, važno je samo da su raspoloživi u tekstualnom obliku i da preporuka na osnovu sličnosti tih opisa ima smisla
* Sistem preporuke koji kombinuje content-based i knowledge-based pristup generisanju preporuka. Konkretno, sistem bi bio zasnovan na [DBpedia](http://wiki.dbpedia.org/) bazi znanja i pristupu opisanom u radu “Linked Open Data for content-based recommender systems”[[5]](#footnote-5). Rad se bavi preporukom filmova, ali opisani pristup kreiranju preporuka je podjednako primenljiv i na druge kategorije stvari (npr. knjige)
* Preporuka Web sadržaja (npr. novinskih članaka ili blog postova) primenom neke od metoda optimizacije, tj. optimizovanog pretraživanja prostora rešenja, kao sto su Hill Climbing, Simulated Annealing, Genetic algorithms.
  + Poglavlje 5 (Optimization) knjige Programming Collective Intelligence[[6]](#footnote-6) koncizno i jasno objašnjava ove metode optimizacije.
  + Detaljniji opis teme, tj šta bi realizacija podrazumevala dat je na sajtu predmeta Intel. Sistemi ([direktan link](http://ai.fon.bg.ac.rs/wp-content/uploads/2015/04/2015-Tema-1-Preporuka-Web-sadrzaja-blog-postova-novinskih-clanaka-1.pdf))
  + Jedan pristup rešavanju datog problema preporuke dat je u [mojim beleškama na ovu temu](https://docs.google.com/document/d/1dyBKM4LgQS94SEGTvTQAf48zeKb4zg8PVY8Qhx2IYho/edit)

Ekstrakcija tema i/ili ključnih reči i izraza

* Aplikacija za efikasnije pretraživanje kolekcije dokumenata primenom prustupa za ekstrakciju ključnih reči / izraza i sumarizaciju teksta zasnovanih na TextRank metodi (opisana u materijalima sa časa). Aplikacija bi omogućila korisniku da pretražuje (npr. kroz neki oblik faceted-browsing-a) dokumente na osnovu ključnih reči / izraza i za izabrani dokument prikazala generisani sumarni prikaz tog dokumenta.
  + dokumenti bi mogli biti novinski članci, blog postovi, opisi filmova, knjiga i sl.
* Prepoznavanje i vizuelizacija ključnih tema u skupu dokumenata. Bilo bi potrebno primeniti (min) dva različita topic modeling pristupa i uporediti rezultate koje oni generišu (na izabranom korpusu tj. skupu dokumenata). Pristupi koji bi se mogli primeniti:
  + LDA (Latent Dirichlet Allocation) - opisan u materijalima sa časa
  + KeyGraph - opisan u materijalima sa časa
  + Non-negative Matrix Factorization, opisan, na primer, u radu “Probabilistic Latent Semantic Analysis” (sekcija 2.2), i lepo ilustrovan u [ovom blog postu](http://tech.opentable.com/2015/01/12/finding-key-themes-from-free-text-reviews/)

Kreiranje prediktivnih modela (ML)

* Kreiranje modela za prepoznavanje astroturfing[[7]](#footnote-7)-a na Twitter-u. Konkretno, potrebno je kreirati prediktivni model (binarni klasifikator) po uzoru na model opisan u radu “Detecting and Tracking Political Abuse in Social Media”[[8]](#footnote-8)
* Prediktivni, klasifikacioni modeli zasnovani na StackExchange podacima (<http://api.stackexchange.com/>); predlozi su dati u okviru [ove teme tj tema za seminarski rad iz Intel. sistema](http://ai.fon.bg.ac.rs/wp-content/uploads/2015/04/2015-Tema-11-Razvoj-prediktivnog-modela-koriscenjem-Stack-Exchange-Datasets.pdf)

1. reviewed for CSEDU2017; clean copy in: /Users/Jelena/Documents/Teaching@FON/MSc Courses [↑](#footnote-ref-1)
2. <http://www.oclc.org/developer/get-started/build.en.html?redirect=true> ; OCLC je najveci svetski katalog knjiga [↑](#footnote-ref-2)
3. <https://www.programmableweb.com/category/books/apis?category=19996> [↑](#footnote-ref-3)
4. D. Blei and J. Lafferty. (2009). Topic models. *Text mining: classification, clustering, and applications*, 10:71. [↑](#footnote-ref-4)
5. <http://sisinflab.poliba.it/semantic-expert-finding/papers/tech-report-1-2012.pdf> [↑](#footnote-ref-5)
6. <http://it-ebooks.info/book/330/> [↑](#footnote-ref-6)
7. Astroturfing is the attempt to create an impression of widespread grassroots support for a policy, individual, or product, where little such support exists; read more at: <https://www.theguardian.com/commentisfree/2012/feb/08/what-is-astroturfing> [↑](#footnote-ref-7)
8. <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2850> [↑](#footnote-ref-8)