UNIVERZITET U BEOGRADU

FAKULTET ORGANIZACIONIH NAUKA

**ZAVRŠNI RAD**

**TEMA:**

Otkrivanje struktura veština za posao pomoću mašinskog učenja i sistema za preporuku

STUDENT: Boris Fidler 2017/3124

Informacini sistemi i tehnologije

Modul: Poslovna inteligencija

Beograd 2018

Komisija koja je pregledala rad kadndidata

Boris Fidler

pod naslovom

Otkrivanje struktura veština za posao

pomoću mašinskog učenja

i sistema za preporuku

i odobrila odbranu:

dr. Miloš Jovanović, docent, mentor

page3image5878864

dr. Milan Vukićević, docent, član komisije

page3image5878864

dr. Sladjan Babarogić, vanredni profesor, član komisije

page3image5878864

# Apstrakt

Ljudski resursi predstavljaju jedan od najbitnijih faktora u jednoj organizaciji. Da bi omogućili da bude zaposlen kvalitetan kadar, neophodno je da svako radno mesto i njegove potrebe budu detaljno analizirani. U ranijim periodima to je bio posao koji se “ručno” radio time što bi zaposleni u odeljenju bili u komunikaciji sa menadžerima odeljenja koja otvaraju pozicije kako bi definisali potrebe novootvorenog radnog mesta. Zatim bi na osnovu primljenih radnih biografija radili selekciju i dalje korake kako bi se došlo do idealnog kandidata. Dolaskom ere kompijuterskih tehnologija i interneta, količina podataka potrebnih za obradu je prevazišla ono što bi se smatralo mogućim za obradu od strane čoveka, neophodno je bilo uključiti i neku vrstu softvera koja bi pomogla u tome. Odeljenje za ljudske resurse se godinama oslanjalo i bilo ograničeno na obradu ponuda od kandidata preko osnovnih menadžerskih aplikacija međutim njih su zamenili ili bolje rečeno njima je pomoglo uključivanje sistema za preporuku.

U ovom radu će biti prikazana analiza sistema za preporuku uz pomoć kojih će proces pravljenja oglasa i regrutacija potencionanih kandidata bili umnogome olakšani.

Uvodni deo rada će opisati proces regrutacije kakav je bio pre Internet ekspanzije i nakon.

Drugi deo rada će biti posvećen opisu trenutnih najvećih sistema koji se bave spajanjem poslodavaca i kandidata.

Rad će se takođe baviti pregledom oblasti iz relevantne literature, prikazom prethodnih i srodnih istraživanja i njihovih zaključaka.

Takođe će biće definisani kriterijumi potrebni za uspešno određivanje veština potrebnih za određeni posao.

Nakon određivanja algoritama i razumevanja podataka na osnovu zadatih kriterijuma biće prezentovan rezultat kao i zaključak sprovedenog istraživanja.

Cilj rada jeste da primenom sistema za preporuku omogući otkrivanje strukture sličnosti između veštine, grupe veština, opisa poslova (u formi slobodnog unosa), i veze između tih veština kako bi se poboljšao proces zapošljavanja kao i proces razvoja karijere za pojedinca. Sistem će dati uvid na sve relevatne veštine za određene poslove i time pomoći poslodavcu da pronađe pravog zaposlenog a onome ko traži posao omogućiti da stekne prave veštine koje bi mu obezbedile posao.

U poslednjem delu ovog rada prikazani su rezultati i benefiti dobijeni projektovanjem ovog modela, kao i zaključci sprovedene analize.

# Abstract

Human resources are one of the most important factors in an organization. In order to enable a high-quality grade of people to be employed, it is necessary that each workplace and its needs be thoroughly analyzed. In earlier periods, it was done as a "manual" job, that would have employees in the department communicate with department managers that open positions, to define the needs of a newly created job opening. Then, on the basis of the received working biographies, the selection and further steps would be made to arrive at the ideal candidate. With the advent of era of computing technologies and the Internet, the amount of data required to process has outperformed what would be considered possible for human processing, it was necessary to include some kind of software to help it. The HR Department has relied on, for years, limited processing of bids from candidates through basic managerial applications, but they have been replaced or, better to say, helped to include a recommendation system.

In this paper will be presented an analysis of the recommendation system by means of which the process of making advertisements and recruitment of potential candidates will be greatly facilitated.

The introductory part of the paper will describe the recruitment process as it was before the Internet expansion and after.

The second part of the paper will be devoted to the description of the current largest systems dealing with the merging of employers and candidates.

The work will also deal with the review of areas from the relevant literature, the presentation of previous and related research and their conclusions.

It will also define the criteria needed to successfully determine the skills required for a specific job.

After determining the algorithms and understanding the data based on the given criteria, the result will be presented as well as the conclusion of the conducted research.

The aim of the paper is to enable the recommendation system to reveal the structure of similarities between skills, skills group, job descriptions (in the form of free entry), and the links between these skills in order to improve the employment process as well as the process of career development for the individual. The system will give insight into all the relevant skills for specific jobs and thus help the employer find the right employee and allow the job seeker to acquire the right skills to secure the job.

The last part of this paper presents the results and benefits obtained by designing this model, as well as the conclusions of the conducted analysis.

Sadrzaj

[Apstrakt 4](#_Toc525305927)

[Abstract 5](#_Toc525305928)

[Prvo poglavlje 8](#_Toc525305929)

[1.1. Uvod 8](#_Toc525305930)

[1.2. Dosadašnji najvažniji rezultati u zadovoljavanju potreba u predmetnoj oblasti 8](#_Toc525305931)

[1.3. Ciljevi i formulacija problema 9](#_Toc525305932)

[1.4. Ciljna grupa i personalna motivacija 9](#_Toc525305933)

[Drugo poglavlje 10](#_Toc525305934)

[2.1. Pregled stanja u predmetnoj oblasti 10](#_Toc525305935)

[2.2. Upravljanje ljudskim resursima 11](#_Toc525305936)

[2.3. Proces regrutacije i selekcija kandidata 11](#_Toc525305937)

[2.4. Uticaj razvoja interneta u načinu oglašavanja poslova 11](#_Toc525305938)

[2.5. Veštine kao deskriptor poslova u IT industriji 11](#_Toc525305939)

[Treće poglavlje 12](#_Toc525305940)

[3.1. Problem istraživanja i metodologija 12](#_Toc525305941)

[3.2. Mašinska obrada prirodnog jezika 12](#_Toc525305942)

[3.3. n-gram 13](#_Toc525305943)

[3.4 Sistemi za preporuku 14](#_Toc525305944)

[3.4.1. Zasnovani na sadržaju 14](#_Toc525305945)

[3.4.2. Zasnovani na saradnji 14](#_Toc525305946)

[3.4.3. Hibridni sistemi 14](#_Toc525305947)

[3.5. TF-IDF 14](#_Toc525305948)

[3.6. Razvoj modela 14](#_Toc525305949)

[Poglavlje četiri 14](#_Toc525305950)

[4.1. Rezultati istraživanja i diskusija 14](#_Toc525305951)

[Poglavlje pet 14](#_Toc525305952)

[5.1. Zakljucak 14](#_Toc525305953)

[Reference 15](#_Toc525305954)

Sistemi za preporuku sve brze postaju defakto nacin za preporuku u razlictim aplikacijama za preporuku prozivoda,usluga i sveukupno informacija korisnicima istih. Mnoge internet aplikacije su se pridruzile trendu koriscenja sistema za preporuku da bi obezbedile veci profit time sto bi korisnicima smanjile vreme pretrage i dolazenja do idealnog proizvoda ili usluge. Neke od takvih kompanija su Amazon, Microsoft,Aliexpress(AliBaba),Netflix i slicno. Sve ove kompanije su uspesno integrisale sisteme preporuke za komercijalnu upotrebu i time uvecale prodaju a samim tim i prihode prilikom prodaje preko interneta i jos bitnije obezbedile lojalnost kupaca. Druge kompanije su takodje razvijale inhouse genericke sisteme preporuke a neki od njih su Net Perceptions,Epiphany, Art Technology, Broad Vision.

Odeljenje za ljudske resurse se godinama oslanjalo i bilo ograniceno na obradu ponuda od kandidata preko osnovnih menadzerskih aplikacija. Ovakvi sistemi je pokretalo interna komunikacija izmedju LJR i drugih odeljenja. Ali sa ekspanzijom kolicine podataka na internetu i uopste podataka u digitalnom formatu i razvojem e-biznisa zahtevalo je odredjenu reformu nacina na koji su kompanije do tog trenutka obavljale proces regrutacije. Platforma za regrutovanje putem interneta jeste jedna od najuspesnijih otkrica putem koje se poslodavci i kandidati otkrivaju. Ovakve platforme su dozivele pravu ekspanziju zbog sve tezeg procesa regrutovanja novih potencijala. Za svaki postavljen oglas za posao, hiljade biografija se svakodnevno moze poslati za otvorene pozicije. Analogno tome takodje postoji ogroman broj radnih biografija koje polako postaju dostupne online. Takava ogromna kolicina dosupnih informacija u vidu poslovnih ponuda, njihovih zahteva kao i radnih biografija je postala odlicno mesto za unapredjivanje kvaliteta u smislu kolicine poklapanja zahteva i ponude. Naravno taj kvalitet je ostao na niskom nivou zbog pristupa logickog operatora (true-false) i time velike kolicine podataka i sansi za dobro poklapanje ostaju neiskorisceni. Iz toga se stvorila potreba za koriscenjem sistema za preporuku kako bi poslodovaci uspesno mogli da obrade ogromne kolicine podataka brzo i efikasno. Svakako resenje nije moguce ostvariti u kratkom roku i samsi sistemi za preporuku predstavljaju izazov u polju istrzivanja koje se i dalje razvija. Kako bi se ta ideja dalje razvijala u ovom radu cemo se baviti nekim problemima i potencionalnim resenjima za unapredjenje sistem za prepruku kod otkrivanja vestina potrebnih za sam posao.

Zapocinjemo sa trenutnim problemima koji prate e-platforme za zaposljavanje i najsavremenijim resenjima koja su predlozena za resavanje poklapanja kandidata tj. Njihovih vestina sa ponudjenim ili bolje receno oglasenim poslovima.

Znacaj informacionih sistema u podrsci sistema za regrutovanje se do sada najvise mogao osetiti u pocetnim fazama kao sto su primanje ponuda od kandidata i obrada njihovih biografija u procesu koji prethodi selekciji. Medjutim sam cin poklapanja izmedju poslova i kandidata je zavisio od promenljivih koje su tesko mogle da budu kvantifikovane. I upravo iz tih razloga informacioni sistemi nisu ucestvovali u kasnijim fazama selekcije. Kao sto je gorenavedeno IS su bili korisceni kao sistem poklapanja u vidu logickog operatora. To znaci da bi do poklapanja dolazilo samo ukoliko bi informacije koje se traze ili nude bile identicne, vestine, tehnologije, iskustvo pre toga i slicno. Ovaj metod je koristio par kljucnih reci za odredjeni posao koje bi se trazile u svim ponudjenim biografijama. Ukoliko bi doslo do poklapanja te biografije bi odlazile na sledecu evaluaciju ali ovaj put direktno od poslodavca bio to rukovodilac odeljenja ili odeljenje za ljudske resurse.

# Prvo poglavlje

## Uvod

Proces “regrutacije” jeste centralna funkcija odeljenja ljudskih resursa pošto upravo ti novi zaposleni postaju faktor u proizvodnji nove vrednosti. Ukratko cilj jeste da proces regrutacije da na svom izlazu novozaposlenog radnika koji će za to preduzeće doneti najveću vrednost.

Sa aspekta poslodavca to podrazumeva pravljenje pozicije koja je okarakterisana nizom zahteva u vidu obrazovanja, kurseva i verovatno najbitnije veština za uspešno obavljanje posla. Od velikog je značaja osigurati potreban broj zaposlenih sa odgovarajućim sposobnostima i kvalifikacijama, kako bi se ostvarili postavljeni ciljevi , zadaci i misije organizacije.

Popunjavanje radnih mesta počinje procesom privlačenja, odnosno regrutovanja ljudskih resursa, a nastavak procesa je odabir, odnosno selekcija kandidata. Da bi se to uradilo pre svega je neophodno otvoriti nove pozicije i u njima definisati potrebne veštine koje kandidat treba da ima kako bi uspešno obavljao posao.

Često imamo situaciju u kojoj naziv pozicije ima dvosmislen naziv ili gore čak da pogresno opisuje potrebe tog posla i veština koje su potrebne za njega. Jedan od načina da se to prevaziđe bi bio da se grupe poslova sortiraju na osnovu naziva. Primenom mašinskog učenja i algoritama takav proces se može unaprediti i ubrzati time što ne bi radio samo analizu naziva pozicija već i veština koje su navedene u okviru njih. Ponavljanje termina među veštinama moze ukazati na one koji najbolje opisuju taj posao ali takođe i da izdvoji opšte od naročitih. Tako odrađena analiza bi u perspektivi mogla da bude aplicirana i na većem skupu podataka sa podjednako dobrim rezultatima.

## 1.2. Dosadašnji najvažniji rezultati u zadovoljavanju potreba u predmetnoj oblasti

## Ciljevi i formulacija problema

Platforme za zapošaljavanje putem interneta polako postaju primarni kanal za oglašavanje i pronalaženje kandidata za većinu kompanija. Dok je takav pristup omogućio da se vreme pronalaženja mogućih kandidata smanji kao i da se značajno umanje troškovi u smislu mesta za oglašavanje, problem tradicionalnih načina je nastavio da postoji.

Pristup logičkog operatora pretrage je jedan od njih gde se sistemom samo direktnog poklapanja dolazilo do određenih kandidata što je prouzrokovalo da mnogi propuste svoju šansu da budu izabrani. Sistemi preporuke imaju za cilj da upravo taj nedostatak nadomeste i da omoguće poslodavcima da dođu do idealnih kandidata. Da bi se to ostvarilo postoje različiti načini kako se uz pomoć sistema za preporuku upravo moze doći do željenih rezultata.

Pronalazenje pravog kandidata “ručno” je dugotrajan i mukotrpan proces, stoga su izmišljene metode koje pomažu u ovom procesu pod nazivom sistema za preporuku za pronalaženje poslova što se tiče mogućih kandidata tj. sistema za regrutovanje sa aspekta poslodavaca.

Odeljenje ljudskih resursa ima zadatak za napravi selekciju kandidata sa odgovarajućim veštinama koje tom preduzeću zaista i trebaju. Koncept veštine je od izuzetnog značaja zato što u mnogim slučajevima može mnogo bolje da oslika potrebu poslodavca ili onoga što moguci kandidat moze da ponudi, nekada i bolje od recimo diplome ili završenog kursa.

Ovaj rad može poslužiti obema stranama u smislu da za poslodavca znači da može da postavi oglas sa najrelevantnijim veštinama i zahtevima za taj posao, dok kandidat može u svakom trenutku da bude upoznat sa nivoom kompleksnosti posla na traženim pozicijama. To znači da bi obe strane profitirale zato što bi sistem za preporuku alalizom strukture traženih veština pomogao potencionalnim kandidatima da steknu prave veštine koje bi im obezbedile posao.

Danas, ako izuzmemo posebno pravljene programe unutar preduzeća iliti “in-house”, društvene mreže poput LinkedIn, Facebook, Twitter, Dice.com i sl. igraju jako bitnu ulogu u procesu regrutovanja zbog informacija koje su dostupne na tim mrežama.

## 1.4. Ciljna grupa i personalna motivacija

U novijem dobu gde je ekspanzijom digitalnih podataka i pojavljivanjem e-platformi za zaposljavanje bila je neohodna reorganizacija nacina na koji kompanije obavaljaju odredjene aktivnosti u odredjenim sferama. Jedna od tih je bila sfera regrutacije. Postavljanja poslovnih ponuda na internet stranicama kompanija se uglavnom vrsilo na delu stranice pod nazivom Karijera (“Carrier”). Zainteresovani bi odlazili na te stranice i aplicirali putem neke online prijave. Medjutim vremenom su se razvile i platforme specijalno dizjanirane za proces regrutacije tj. Zaposaljvanja. Na tim mestima su zainteresovani mogli da naprave svoj profil koji bi zatim popunili sa kljucnim informacijima u vezi sa njihovim skolovanje, prethodnim iskustvom i vestinama koje poseduju. Nakon toga bi ukoliko bi se otvorila nova pozicija imali mogucnost da apliciraju na istu i time steknu sansu u buducem zaposlenju. Nazalost ovakav sistem je bio odrziv samo u ranijim fazama interneta. Trenutna situacija je bila takava da bi za odredjenu poslovnu ponudu stizale hiljade odgovora tj apliciranja. Iz ovoga se stvorila potreba za sistemima za preporuku.

Sa druge strane cak i trazenje posla je nekada mogao biti iscrpljujuca aktivnost. Najcesci pristup bi bio da se uz pomoc par kljucnih reci izvrsi pretraga na stranici potencionalnog poslodavca. Rezultat pretrage bi vratio listu poslova koji sadrze neku od tih reci. Nazalost takav rezultat nije garantovao da bi kandidatu odgovarao posao na osnovu njegovih sklonosti i znanja.

# Drugo poglavlje

## 2.1. Pregled stanja u predmetnoj oblasti

Odeljenje za ljudske resurse se godinama oslanjalo i bilo ograniceno na obradu ponuda od kandidata preko osnovnih menadzerskih aplikacija. Ovakvi sistemi je pokretalo interna komunikacija izmedju LJR i drugih odeljenja.

Ali sa ekspanzijom kolicine podataka na internetu i uopste podataka u digitalnom formatu i razvojem e-biznisa zahtevalo je odredjenu reformu nacina na koji su kompanije do tog trenutka obavljale proces regrutacije. Platforma za regrutovanje putem interneta jeste jedna od najuspesnijih otkrica putem koje se poslodavci i kandidati otkrivaju. Ovakve platforme su dozivele pravu ekspanziju zbog sve tezeg procesa regrutovanja novih potencijala. Za svaki postavljen oglas za posao, hiljade biografija se svakodnevno moze poslati za otvorene pozicije. Analogno tome takodje postoji ogroman broj radnih biografija koje polako postaju dostupne online.

Takava ogromna kolicina dosupnih informacija u vidu poslovnih ponuda, njihovih zahteva kao i radnih biografija je postala odlicno mesto za unapredjivanje kvaliteta u smislu kolicine poklapanja zahteva i ponude. Naravno taj kvalitet je ostao na niskom nivou zbog pristupa logickog operatora (true-false) i time velike kolicine podataka i sansi za dobro poklapanje ostaju neiskorisceni. Iz toga se stvorila potreba za koriscenjem sistema za preporuku kako bi poslodovaci uspesno mogli da obrade ogromne kolicine podataka brzo i efikasno. Svakako resenje nije moguce ostvariti u kratkom roku i samsi sistemi za preporuku predstavljaju izazov u polju istrzivanja koje se i dalje razvija.

Kako bi se ta ideja dalje razvijala u ovom radu cemo se baviti nekim problemima i potencionalnim resenjima za unapredjenje sistem za prepruku kod otkrivanja vestina potrebnih za sam posao.

## 2.2. Upravljanje ljudskim resursima

Svaka organizacija ima potrebu za ljudima, ali i ljudi imaju porebu da budu deo oragnizacije, kako bi zajedničkim delovanjem ostvarili zadate ciljeve poslovanja. Svaka strategija poslovanja, neminovno polazi od ljudskog kapitala, koji ima najvažniju ulogu u procesu rada. Kako tehničko-tehnološke, ali i društvene promene, zahtevaju različite profile kadrova, nivo zahtevanog obrazovanja zaposlenih se povećava, a s druge strane, sve veća podela rada uslovljava stručne specijalizacije, što dovodi do ograničene mobilnosti ljudskih resursa. Planiranje rasta i razvoja organiazcije mora da bude prožeto nastojanjima da se obezbedi potrebna strukutura zaposlenih, dok su ljudski resursi upravo i inicijator rasta i ostalih promena u poslovanju organizacije.

Kako su kadrovi osnovi stub svakog poslovanja, njihovo planiranje mora da bude deo strategije rasta organizacije, a kako bi se ostavrila veza između planiranja razvoja kadrova i poslovne strategije, moraju da postoje osnovni nivoi planiranja: (Petković M. J. N., 2005)

Aktivnosti sektora za ljudske resurse su: planiranje ponude i tražnje za ljudskim resurisma, analkiza posla, regrutovanje potencijalnih kandidata, selekcija kandidata i njihova socijalizacija po prijemu na rad, ocenijivanje performansi zaposlenih, njihova obuka i razvoj karijere, nagrađivanje, radni odnosi i kolektivno pregovaranje, upravljanje procesom napuštanja organizacije, zdravlje i bezbednost zaposlenih na radu i ostalo.

## 2.3. Proces regrutacije i selekcija kandidata

## 2.4. Uticaj razvoja interneta u načinu oglašavanja poslova

## 2.5. Veštine kao deskriptor poslova u IT industriji

# Treće poglavlje

## 3.1. Problem istraživanja i metodologija

## 3.2. Mašinska obrada prirodnog jezika

Svakodevno, ljudi razmenjuju hiljade reci koje drugi ljudi mogu da interpretiraju na razlicite nacine. Jednostavnije receno radi se o komunikaciji, ali naravno uvek treba imati na umu da reci mogu imati mnogo dublje znacenje u zavisnosti od konteksta. Masinska obrada prirodnog jezika (Natural Language Processing - NLP) se fokusira na konteksutalni patern pre nego na vokalni nacin na koje su reci izgovorene (Mills, 2018)

Masinska obrada prirodnog jezika je u svojoj osnovi forma vestacke inteligencije koja analizira ljudski jezik. Postoje varijateti medjutim sve imaju zajednicku osobinu da predstavljaju tehnologiju koja pomaze masinama da razumeju nas jezik, pa cak i da komunicijaju sa nama koristeci nas jezik.

Svoju istoriju je zapocela 1950 godine kao mesavina vestacke inteligencije i ligvistike. Njegovi koreni vuku iz tehnologije za izvlacenje teksta (Informational retreval - IR), koja primenjuje statisticki zasnovane tehnike da indeksira i pretrazi velike kolicine teksta efikasno.

Odlike prirodnog jezika kao sto su velicina, nestruktuirana priroda kao i cinjenica da je podlozno interpretaciji dovelo je do problema kada su u pitanju standardni pristupi parsiranju.

Masinska obrada prirodnog jezika mora da omoguci izvlacenje sematnike iz teksta uzimajuci u obzir delove govora tj. teksta kao sto su imenice, glagoli, pridevi i sintaksa recenice. (Prakash, Lucila, & Wendy, 2011)

Recnik se sastoji od Tekst cini niz reci Jezik je konstruiran

skupa reci iz recnika iz skupa svih mogucih tekstova



(Neves, 2016)

## 3.3. n-gram

Kategorizacija teksta predstavlja kljucni zadatak u obradi dokumenata, cime se stice mogucnost obrade ogromne kolicine podataka koji se nalaze u elektronskoj formi. (Trenkle, 2001) Ono sto predstavlja realnu poteskocu u obradi teksta u elektronskim dokumentima jesu tekstualne greske bile ona gramatickog ili sintaksnog tipa. Da bi sistem teksutalne kategorizacije bio proglasen pouzdanim sva razlicite vrste izvoda podataka neophodno je da ipak ima neku dozu tolerancije na oredjene greske kao i nacin da ih prevazidje.

Dokumenta u elektronskom formatu poticnu iz raznih izvora. Neki su generisani kao deo programa koji je zaduzen za pisanje teksta i kao takvi podredjeni su poslovnoj logici ili tzv. Speel-check programa dok su drugi slobodnog formata unosa poput email-a. Druga pomenuta grupa je uglavnom ona koja se stvara na licu mesta bez neke preprovere ili provere nakon stvaranja, kao sto su skenirani dokumenti uz pomoc posebnih softvera za prebacivanje iz papirne u elektronsku formu.

Prilikom skeniranja ili bolje receno prepoznavanja teksta sa papira neminovno je da ce nastati neka greska i upravo je to tip situacije koji bi zahtevao dalju proveru koja je skupa i komplikovana i gde bi neki sistem za proveru teksta bio dobrodosao.

U ljudskom jeziku neke reci se pojavljuje cesce od drugih. Jedan od osnovnih nacina da se iskaze takva ideja je danas poznata kao Zip-ov zakon (Kingsley, 1950). On kaze da je pojavljivanje n-te najcesce reci u ljudskom jeziku u tekstu inverzno proporcionalno n. Ovo znaci da najcesca rec u nekom jeziku se pojavljuje dva puta cesce od druge reci i tri puta cesce od trece itd, implicirajuci da uvek postoje reci koje su na neki nacin dominatne u tom jeziku. Ovakav zakljucak se odnosi i na reci generalno ali i na oredjene teme.

n-gram predstavlja kontinuiranu sekvencu od n reci iz neke recenice ili teksta na osnovu govora. To znaci da to moze biti veznik, slovo, rece, slozenica ali ono sto ih sve odlikuje jeste da se posmatraju kao atomska vrednost. N-gram od jedne “stavke” tj. reci se naziva “unigram”,od dve “bigram”,tri “trigram” itd. (Hong, Nduyen, Duong, & Snasel, 2016)

Ovakva podela ce biti od izuzetne koristi u ovom radu posto se u pretprocesiranju radi obrada teksutalnih polja za vestine kao i za sam opis posla

## 3.4 Sistemi za preporuku

Sistemi za preporuku sve brze postaju defakto nacin za preporuku u razlictim aplikacijama za preporuku prozivoda,usluga i sveukupno informacija korisnicima istih. Mnoge internet aplikacije su se pridruzile trendu koriscenja sistema za preporuku da bi obezbedile veci profit time sto bi korisnicima smanjile vreme pretrage i dolazenja do idealnog proizvoda ili usluge. Neke od takvih kompanija su Amazon, Microsoft,Aliexpress(AliBaba),Netflix i slicno. Sve ove kompanije su uspesno integrisale sisteme preporuke za komercijalnu upotrebu i time uvecale prodaju a samim tim i prihode prilikom prodaje preko interneta i jos bitnije obezbedile lojalnost kupaca. Druge kompanije su takodje razvijale inhouse genericke sisteme preporuke a neki od njih su Net Perceptions,Epiphany, Art Technology, Broad Vision.

### 3.4.1. Zasnovani na sadržaju

### 3.4.2. Zasnovani na saradnji

### 3.4.3. Hibridni sistemi

## 3.5. TF-IDF

## 3.6. Razvoj modela

# Poglavlje četiri

## 4.1. Rezultati istraživanja i diskusija

# Poglavlje pet

## 5.1. Zakljucak

# **Reference**

Hong, N. V., Nduyen, H., Duong, H. N., & Snasel, V. (2016). n -Gram-Based Text Compression. *Computational Intelligence and Neuroscience*.

Kingsley, G. (1950). *Human begavior and the principle of least effort: An introduction to human ecology.* Cambrige.

Mills, T. (2018, Jul 2). *What Is Natural Language Processing And What Is It Used For?* Retrieved from Forbes: https://www.forbes.com/sites/forbestechcouncil/2018/07/02/what-is-natural-language-processing-and-what-is-it-used-for/#7ee74645d71f

Neves, M. (2016, April 11). Retrieved from Semanticscholar: https://pdfs.semanticscholar.org/presentation/a575/e0cef057241668a53e75f0627189a0e7e92a.pdf

Prakash, N. M., Lucila, O.-M., & Wendy, C. W. (2011, September). *Natural language processing: an introduction.* Retrieved from https://dx.doi.org/10.1136%2Famiajnl-2011-000464

Trenkle, W. B. (2001). N-Gram-Based Text Categorization. *ResearchGate*. Retrieved from http://odur.let.rug.nl/vannoord/TextCat/textcat.pdf