

KOMISIJI ZA NASTAVU II STEPENA ELEKTROTEHNIČKOG FAKULTETA U BEOGRADU

Komisija za nastavu II stepena Elektrotehničkog fakulteta u Beogradu, imenovala nas je za članove komisije za pregled i ocenu master rada kandidata Nikole Miloševića, sa brojem indeksa 11/3138, pod naslovom „Mašinska analiza sentimenta rečenica na srpskom jezika“. Komisija je pregledala master rad i podnosi sledeći izveštaj:

IZVEŠTAJ

I Biografski podaci o kandidatu

Nikola Milošević je rođen 07. 12. 1986. godine u Bratislavi, Slovačka. Osnovnu i srednju školu je završio u Beogradu sa odličnim uspehom.

Školske 2006/2007 godine je upisao Elektrotehnički fakultet, Univerziteta u Beogradu, smer Računarska tehnika i informatika. Diplomirao je 29. 12. 2010. godine na temu „Izvoz testa prema šablonu iz programa testBase za program MS Word“.

Školske 2011/2012 godine je upisao master akademske studije na Elektrotehničkom fakultetu, Univerziteta u Beogradu, smer računarska tehnika i informatika.

U decembru 2009. godine dobio je priznanje za uspešan razvoj programskih dodataka za program Internet Explorer 8, za web portal fakulteta od kompanije Microsoft Srbija. Dobitnik je druge nagrade na takmičenju Cyber security iPuzzle u pronalaženju bezbednosnih propusta u web aplikacijama, koje je organizovano u okviru Međunarodnog sajma bezbednosti objekata, infrastrukture, lica i poslovanja (iSEC) 2011. godine.

II Osnovni podaci o radu

U okviru rada je razvijen mašinski analizator sentimenta za srpski jezik. Urađeni softver ima tri dela: stemer za srpski jezik, koji na osnovu relativno malog broja pravila uklanja sufikse reči i na taj način olakšava mašinsko učenje, softver za unošenje trening skupa labelisanih rečenica na osnovu kojih uz algoritam supervizovanog mašinskog učenja uči karakteristike klasa i analizator sentimenta, koji na osnovu naučenih podataka binarno klasifikuje uneti tekst na dve klase - pozitivnu i negativnu. Algoritam mašinskog učenja, upotrebljen za klasifikaciju teksta, je Naive Bayes. Softver je realizovan u vidu web servisa, sa kojim se može komunicirati uz pomoć HTTP POST zahteva koji sadrže JSON (JavaScript Object Notation) objekata.

Analiza sentimenta se već nekoliko godina koristi za razne namene (istraživanje tržišta i javnog mnjenja, marketing, eZakonodavstvo...). Analizom postojećih rešenja primećeno je da različita rešenja procesuiranja jezika, a među njima i analize sentimenta postoje za mnoge jezike, ali rešenja za srpski jezik su retka. Predloženi sistem se bavi upravo klasifikacijom teksta za srpski jezik, kao i stemmingom, koji predstavlja radnju pripreme teksta za druge zadatke obrade jezika (pribavljanje informacija, odgovaranje na pitanja, klasifikacija, mašinsko prevođenje...). U radu je dat istorijski pregled razvoja oblasti, predloženo rešenje za srpski jezik i opisana implementacija sistema.

Master rad ima šest glava, od kojih su prva i šesta redom uvod i zaključak, literaturu i prilog.

III Analiza rada

U prvoj uvodnoj glavi dat je pregled oblasti analize sentimenta, kao i trenutnih rešenja koji koriste analizu sentimenta poput Google Product Search ili Bing Shopping. Izložene su

nove tendencije u svetu analize sentimenta, dati statistički podaci osnovnih algoritama, formulisan problem i predloženo rešenje.

U drugoj glavi predočeni su motivi i detaljno definisani zahtevi rada. Rad rešava dva problema - problem otklanjanja sufiksa radi bržeg učenja prilikom korišćenja algoritama mašinskog učenja i problem klasifikacije teksta, odnosno specijalni slučaj klasifikacije teksta, kakva je analiza sentimenta.

U trećoj glavi je opisan stemming, odnosno algoritam koji uz pomoć ograničenog broja pravila uklanja sufikse reči u različitim oblicima i formira jedinstveni oblik reči za sve gramatičke oblike, nalik osnovi. Dat je istorijski pregled razvoja stemming-a za različite jezike, analizirana su predhodna rešenja za srpski jezik, kao i za srodne jezike, opisan način formiranja predloženog stemera, data evaluacija predloženog stemera i opisani njegovi nedostaci.

U četvrtoj glavi je dat istorijski pregled razvoja analize sentimenta, opisani različiti slučajevi korišćenja analize sentimenta, opisan način rada algoritma Naive Bayes, koji je upotrebljen u implementaciji rešenja, opisane neke njegove mane i dati predlozi za prevazilaženje nekih od mana algoritma.

U petoj glavi je prikazan način na koji je sistem realizovan. Objasnjena je arhitektura realizovanog softvera, model baze podataka, kao i način na koji se realizovani web servis može koristiti.

U šestoj glavi dat je zaključak u kojem se rezimiraju rezultati rada. Navedeni su neki nedostaci postojećeg rešenja kao i ideje za moguću nadogradnju.

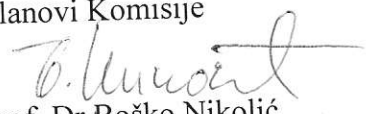
IV Zaključak i predlog

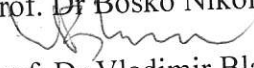
Kandidat Nikola Milošević je, na osnovu analize postojećih softverskih alata iz oblasti mašinske obrade jezika i zahteva koje bi realizovani alat morao da zadovolji, najpre predložio odgovarajuće rešenje, a zatim dizajnirao sistem za analizu sentimenta rečenica na srpskom jeziku, kompletno realizovao pomenuti sistem i na kraju dao kritički osvrt na dobijeno rešenje. Razvijeni web servis omogućava analizu sentimenta rečenica na srpskom jeziku, odnosno klasifikuje zadati tekst u pozitivnu ili negativnu klasu. Takođe, web servis se može upotrebiti samo za stemming prilikom neke druge obrade teksta, koja kao pripremu zahteva stemming obradu. Koncipiranjem i realizacijom ovog sistema kandidat je doprineo jednostavnijem i kvalitetnijem radu u svetu mašinske obrade jezika i klasifikacije teksta.

Na osnovu svega izloženog članovi Komisije predlažu Nastavno-naučnom veću Elektrotehničkog fakulteta u Beogradu da prihvati master rad kandidata Nikole Miloševića, diplomiranog inženjera elektrotehnike, pod naslovom „Mašinska analiza sentimenta rečenica na srpskom jeziku“ i odobri njegovu javnu odbranu.

U Beogradu, 18.12.2012. godine

Članovi Komisije


Prof. Dr. Boško Nikolić


Prof. Dr. Vladimir Blagojević