

KOMISIJI ZA STUDIJE II STEPENA ELEKTROTEHNIČKOG FAKULTETA U BEOGRADU

Komisija za studije II stepena, Elektrotehničkog fakulteta u Beogradu, na svojoj sednici održanoj 04.06.2013. godine imenovala nas je u Komisiju za pregled i ocenu master rada dipl. inž. Aleksandra Matića pod naslovom „Distribuirano prikupljanje podataka sa veba“. Nakon pregleda materijala Komisija podnosi sledeći

IZVEŠTAJ

1. Biografski podaci kandidata

Aleksandar B. Matić je rođen 03.08.1980. godine u Loznici. Gimnaziju je završio u Loznici sa odličnim uspehom. Elektrotehnički fakultet u Beogradu upisao je 1999. godine, na odseku za Računarsku tehniku i informatiku. Diplomirao je u septembru 2007. godine sa prosečnom ocenom na ispitima 7.43, na diplomskom 10. Master studije na Elektrotehničom fakultetu u Beogradu je upisao školske 2010-2011 na odseku za računarsku tehniku i informatiku. Položio je sve ispite sa prosečnom ocenom 8.33.

2. Opis master rada

Master rad kandidata sadrži 51 stranu teksta, zajedno sa slikama i dodacima. Rad sadrži 16 delova, spisak referenci i spisak literature. Spisak referenci sadrži 15 referenci dok spisak literature sadrži 2 knjige koje detaljno objašnjavaju alate korišćene u radu.

Prva dva dela u radu uvode čitaoca u problematiku kojom se rad bavi i objašnjavaju motive i ciljeve u radu. Dalje, u trećem delu rada objašnjena je osnovna terminologija i problemi u implementaciji kako *full* tekst pretrage tako i fokusirane pretrage koja je i glavna tema ovog rada. Četvrti deo je deo u kom se čitalac upoznaje sa široko rasprostranjenim algoritmima koji se koriste za prikupljanje podataka sa interneta.

Peti deo rada predstavlja Map Reduce paradigmu kroz tipove koji se koriste i samu implementaciju. Map Reduce predstavlja osnovu koja omogućava distribuiranu pretragu tako da su terminologija kao i sam algoritam objašnjeni sa nešto više detalja. Ovo predstavlja uvod u šesti deo rada koji upoznaje čitaoca sa Hadoop platformom za procesiranje podataka. Hadoop je platforma na kojoj se izvršava ranije spomenuti Map Reduce.

U sedmom delu dolazimo do prvih konkretnih tema u vezi sa distribuiranim prikupljanjem podataka sa veba. Sedmi deo zapravo povezuje Hadoop platformu (za distribuirano procesiranje podataka) na kojoj se izvršava Map Reduce i Nutch biblioteku koja predstavlja alat koji omogućava relativno jednostavno prikupljanje i full tekst pretragu podataka sa interneta.

Delovi 8,9,10 i 11 objašnjavaju sa nešto više detalja Nutch, strukture podataka u njemu, kao i kakve mogućnosti ima za proširivanje funkcionalnosti. Takođe tu su objašnjene i metode kojima

je moguće pokrenuti Nutch pretragu.

Algoritam koji omogućava vertikalnu pretragu podataka pomoću navedenih alata je predstavljen u 12 delu rada. Potom u delovima 13 i 14 slede implementacija kao i rezultati testova koji su pokrenuti kako bi se prikazale mogućnosti algoritma. Poslednja 3 poglavlja su zaključak i spisak literature.

3. Analiza rada sa ključnim rezultatima

U master radu dipl. inž. Aleksandra Matića realizovan je projekat pod nazivom "Distribuirano prikupljanje podataka sa veba" u kome je nad distribuiranom platformom za prikupljanje podataka (Nutch) implementiran algoritam koji omogućava vertikalnu pretragu.

Osnovni doprinosi rada su:

- a) Realizovana je implementacija algoritma ponuđenog u radu. Kod je javno dostupan.
- b) Uspešno je demonstrirana integracija Nutch platforme pri rešavanju konkretnog problema.
- c) Demonstrirana je mogućnost da se čitav logički Hadoop klaster izvršava na jednoj fizičkoj mašini
- d) Demonstrirana je mogućnost proširivanja funkcionalnosti Nutch platforme na manje uobičajen način, odnosno direktnim implementiranjem MapReduce paradigme.

4. Zaključak i predlog


Kandidat Aleksandar Matić je u svom master radu uspešno rešio problem distribuirane vertikalne pretrage podataka uz pomoć Nutch platforme. Predloženo rešenje može uspešno da se koristi za vertikalnu pretragu iako Nutch nije inicijalno zamišljen za taj tip pretrage.

Kandidat je u izradi rada pokazao samostalnost, sistematičnost i temeljitost. On je kroz istraživanje izvršeno tokom izrade rada značajno unapredio svoje znanje iz oblasti distribuirane obrade podataka kroz Map Reduce paradigmu.

Na osnovu gore navedenog Komisija predlaže Nastavno-naučnom veću Elektrotehničkog fakulteta u Beogradu da prihvati rad „Distribuirano prikupljanje podataka sa veba“ dipl. inž. Aleksandra Matića kao master rad i odobri javnu usmenu odbranu.

Beograd 12.06.2013.

Članovi komisije:


dr. Veljko Milutinović


dr. Miroslav Bojović