

НАСТАВНО-НАУЧНОМ ВЕЋУ

Предмет: Подобност теме и кандидата Милоша Павковића, мастер инжењера електротехнике и рачунарства за израду докторске дисертације

Одлуком Наставно-научног већа Електротехничког факултета бр. 5002/11-1 од 13.12.2017. године, именовани смо за чланове Комисије за оцену подобности теме и кандидата Милоша Павковића, мастер инжењера електротехнике и рачунарства, за израду докторске дисертације и научне заснованости теме **“Побољшање перформанси прикупљања кориснички генерисаних садржаја на Вебу применом адаптивних интелигентних метода”**.

На основу материјала приложеног уз захтев кандидата, Комисија подноси следећи

ИЗВЕШТАЈ

1. Подаци о кандидату

1.1. Биографски подаци

Милош Павковић је рођен 14. новембра 1982. у Београду. Основну школу “Душан Јерковић” и техничку школу “Михајло Пупин” је завршио у Инђији. У октобру 2004. године уписао је основне студије на Електротехничком факултету, Одсек за софтверско инжењерство. Дипломирао је 2009. године, са просечном оценом 9, а дипломски рад је одбранио са оценом 10. Током студија, у оквиру “*Best Student Recognition*” догађаја, изабран је од стране IBM компаније за учешће у “*Internship*” програму, где је радио на пројекту “*A comparison of portlet container in IBM WebSphere Application Server and IBM WebSphere Portal*”. Током пројекта боравио је у научном-истраживачком центру IBM-а у Ници, Француска.

Након дипломирања уписао је мастер студије на Електротехничком факултету 2009. године, на модулу Софтверско инжењерство. Положио је све испите предвиђене планом и програмом са оценом 9.83 и одбранио мастер рад са темом “Софтверски системи за примену Бајесових мрежа у медицини” са оценом 10.

На Електротехничком факултету је 2011. године уписао докторске студије на модулу Софтверско инжењерство код ментора проф. др Јелице Протић и положио све предмете са просечном оценом 10. У свом истраживачком раду показао је посебно интересовање за интелигентне системе претраживања на интернету, паралелизацију претраживања Веб форума, аутоматску екстракцију података и примену машинског учења на аутоматизацију генерисања регуларних израза. Конципирао је и развио специјализовани Веб форум претраживач, који кластеризацијом Веб страна и применом машинског учења аутоматски прати скелетне линкове Веб форума на оптималан начин и врши екстракцију кориснички генерисаног садржаја. Резултате тестова и истраживања је објавио на две међународне и три домаће конференције.

За време докторских студија, 10. октобра 2013. године изабран је у звање асистента за ужу научну област Рачунарске комуникације на Природно-математичком факултету Универзитета у Крагујевцу, и поново изабран у исто звање 10. октобра 2016. године. Као

асистент на Природно-математичком факултету ангажован је на вежбама из предмета: Архитектура рачунара 1, Рачунарске мреже и мрежне комуникације, Оперативни системи 1. Додатно је ангажован од октобра 2017. године као асистент на Факултету инжењерских наука Универзитета у Крагујевцу, где држи предмет Рачунарске основе интернета на модулу за Софтверско инжењерство.

Од 1. новембра 2010. године па до данас је ангажован од стране фирме *Effyis Inc*, са седиштем у Детроиту, Сједињене Америчке Државе, на конципирању и изради специјализованих претраживача Веба.

У септембру 2006. године је основао фирму *PulsArt*. Током рада за ову фирму учествовао је у пројектовању и имплементацији локалних рачунарских мрежа, пројектовању инфраструктуре за паметне зграде, информационих система, Веб портала, ЦМС Веб базираних система и Веб сајтова.

1.2. Стечено научноистраживачко искуство

Милош Павковић је научно-истраживачко искуство започео уписом на докторске академске студије на Електротехничком факултету Универзитета у Београду школске 2011/2012 године, на модулу Софтверско инжењерство, и на тим студијама положио је следеће испите:

- Развој микропроцесорског софтвера (9 ЕСПБ)	оцена 10 (десет)
- Виртуелна стварност (9 ЕСПБ)	оцена 10 (десет)
- Медицинска информатика (9 ЕСПБ)	оцена 10 (десет)
- Индексирање и претраживање мултимедијалних садржаја (9 ЕСПБ)	оцена 10 (десет)
- Енглески језик (6 ЕСПБ)	оцена 10 (десет)
- Софтверска техника (9 ЕСПБ)	оцена 10 (десет)
- Сигурност и заштита рачунарских система (9 ЕСПБ)	оцена 10 (десет)
- Вештачка интелигенција и експертски системи (9 ЕСПБ)	оцена 10 (десет)
- Интернет програмирање (9 ЕСПБ)	оцена 10 (десет)
- Моделовање и мерење рачунарских перформанси (6 ЕСПБ)	оцена 10 (десет)

Поред испита, кандидат је обавио и два студијска истраживачка рада (15+15 ЕСПБ) и један научно-стручни рад (3 ЕСПБ), што је предвиђено планом докторских студија. Тиме је кандидат остварио укупно 120 ЕСПБ.

Милош Павковић је аутор 6 научних публикација. Аутор је три рада на међународним конференцијама, и три рада на домаћим конференцијама.

Главне област интересовања су му интелигентни системи претраживања на интернету, паралелизација претраживања Веб форума, аутоматска екстракција података и примена машинског учења на аутоматизацију генерисања регуларних израза.

Дана 26.12.2017. године одржана је јавна усмена одбрана теме докторске дисертације Милоша Павковића, пред комисијом коју су чинили:

- др Бошко Николић, редовни професор Електротехничког факултета Универзитета у Београду,
- др Милош Ковачевић, ванредни професор Грађевинског факултета Универзитета у Београду, и
- др Марија Рашајски, ванредни професор Електротехничког факултета Универзитета у Београду.

Комисија је кандидата на јавној усменој одбрани оценила оценом задовољно.

М33 – међународне конференције:

- M. Pavković, D. Drašković, L. Šubelj, S. Žitnik, D. Lavbič, M. Janković, J. Protić, B. Nikolić “Intelligent techniques for searching Internet forums”, *Proceedings of the 22nd International Electrotechnical and Computer Science Conference ERK 2013*, 16-18. September, 2013, Portorož, Slovenia, pp. B85-B88.
- M. Pavković, J. Protić, “The use of an intelligent forum crawler for data retrieval from e-learning portals”, *Proceedings of the 6th International Conference on Education and New Learning Technologies*, 7-9. July, 2014, Barcelona, Spain, pp. 2441-2449
- M. Pavković, J. Protić, “An Analysis of Scientific Publications from Serbia: The Case of Computer Science”, *Proceedings of the 15th International Conference on Scientometrics and Informetrics*, 29. June – 4. July, 2015, Istanbul, Turkey, pp. 423 – 424.

М63 – домаће конференције:

- D. Drašković, M. Pavković, “Inteligentne tehnike za pretraživanje Interneta”, 57. *Konferencija za elektroniku, telekomunikacije, računarstvo, automatiku i nuklearnu tehniku ETRAN*, Zlatibor, Srbija, 3-6. Jun, 2013, RT2.9.
- M. Pavković, J. Protić, “Intelligent Crawler for Web Forums based on Improved Regular Expressions”, *Proceedings of the 21st Telecommunications forum TELFOR*, Belgrade, Serbia, November 26-28, 2013, pp. 817-820.
- M. Pavković, J. Protić, “Paralelni pretraživač za indeksiranje stranica foruma na Internetu”, 58. *Konferencija za elektroniku, telekomunikacije, računarstvo, automatiku i nuklearnu tehniku ETRAN*, Vrnjačka Banja, Srbija, November 2-5. Jun, 2014, RT5.4.

1.3. Оцена подобности кандидата за рад на предложеној теми

Приликом писања овог Извештаја, Комисија је размотрила све претходно наведене резултате досадашњег рада кандидата Милоша Павковића. Узимајући у обзир претходно објављене научне радове, положене испите, стручно усавршавање у иностранству, обављене стручне обавезе током докторских студија и успешну одбрану теме докторске дисертације може се закључити да кандидат има знање и способности неопходне за рад на предложеној теми. Комисија сматра да је кандидат способан да током израде докторске дисертације оствари резултате који ће бити верификовани од стране међународне научне заједнице. Став Комисије је позитиван и предлаже се да тема буде усвојена, чиме се омогућава да кандидат Милош Павковић, који је испунио све потребне и формалне услове, приступи изради докторске дисертације под насловом “Побољшање перформанси прикупљања кориснички генерисаних садржаја на Вебу применом адаптивних интелигентних метода“.

2. Предмет и циљ истраживања

Платформе за дискусије на интернету, као што су форуми, “Q&A” сајтови и сајтови са корисничким коментарима, данас представљају значајан део друштвених интернет страница. Надлазеће друштвене и медијске мреже су измениле првобитну намену оваквих сајтова који су еволуирали у специјализоване алате за размену мишљења експерата. Дискусионе платформе данас представљају најбољи медиј за многе стручне размене информација о темама и питањима од општег интереса. Прикупљање и анализа кориснички генерисаног садржаја ових дискусија обезбеђује вредне корисничке податке, као и сакупљање и систематизацију знања и испитивање јавног мњења.

На дискусионим платформама кориснички генерисан садржај се скоро никада не мења нити брише, него се стално додаје. Из тог разлога је неопходан ефикасан инкрементални претраживач који проналази и прикупља искључиво нове садржаје, како би пропусни опсег, време прикупљања података и укупне перформансе система претраживања биле оптималне. Веб форум као један од најзаступљенијих облика дискусионе платформе, веома је погодан за анализу и примену разних техника претраживања и прикупљања информација. Проблем са генеричким претраживачима је тај што имају тенденцију да третирају сваку страницу Веб сајта као појединачни објекат који посећују повремено, мењајући стари садржај базе података са новим прикупљеним садржајем. Овај приступ није адекватан за дату структуру платформе за дискусију и презентацију садржаја. Са друге стране, специјализовани форум претраживачи, као што су FoCUS и iRobot, су дизајнирани за преузимање само важних страница форума праћењем искључиво скелетних УРЛ-ова Веб форума. Остављајући остале неважне странице (као што су стране за пријављивање, претраживање, корисничке или странице за помоћ итд.), ови претраживачи су ефикаснији у смислу пропусног опсега и покривености од генеричког претраживача приликом претраживања Веб форума по први пут. Примена генеричког претраживача доводи до поновног скидања целокупног садржаја који је већ прикупљен у претходним посетама, услед немогућности препознавања и циљања новонасталог садржаја између две посете.

Циљ овог рада је конципирање и развијање инкременталног Веб претраживача, примењеног на дискусионе платформе, а посебно Веб форуме, који кластеризацијом Веб страна и применом машинског учења аутоматски прати скелетне линкове Веб форума, врши екстракцију кориснички генерисаног садржаја и интелигентно се адаптира тренутној технологији тако да време, проток и ресурси рачунара буду оптимално утрошени. Додатно, рад ће дати преглед метода и модула у циљу ефикасног прикупљања искључиво новог садржаја у инкременталној стратегији претраживања форума. Ефикасност би била постигнута откривањем и коришћењем карактеристика Веб сајт технологије, затим начином на који је садржај представљен као и доступним навигационим опцијама између страница. Имплементација предложених решења би омогућила претраживачу да дође до новог садржаја на најефикаснији начин, без обиласка страница на којима се налази већ преузет садржај. Директна корист оваквог приступа је: а) побољшана искоришћеност пропусног опсега и времена претраге за новим садржајем; б) оптимизована упосленост ресурса рачунара на којима се претраживачи извршавају; в) спречавањем прикупља редундантних података се повећава квалитет постојећих.

Референце истраживања везаних за тезу:

- [1] Cho, J. and H. Garcia-Molina, *The Evolution of the Web and Implications for an Incremental Crawler*, in Proceedings of the 26th International Conference on Very Large Data Bases. 2000, Morgan Kaufmann Publishers Inc. p. 200-209.
- [2] Cai, R., et al., *iRobot: an intelligent crawler for web forums*, in Proceedings of the 17th international conference on World Wide Web. 2008, ACM: Beijing, China. p. 447-456.
- [3] Jiang, J., et al., *FoCUS: Learning to Crawl Web Forums*. IEEE Transactions on Knowledge and Data Engineering, 2013. 25(6): p. 1293-1306.
- [4] Pavković, M. and J. Protić. *The parallel crawler for indexing forum pages on the Internet*. in 2014 21st Conference on Electrical, Electronic and Computing Engineering (ETRAN). 2014. Vrnjacka Banja, Serbia.
- [5] Yang, J.-M., et al., *Incorporating site-level knowledge for incremental crawling of web forums: a list-wise strategy*, in Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. 2009, ACM: Paris, France. p. 1375-1384.

- [6] Wang, Y., et al., *Exploring traversal strategy for web forum crawling*, in Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval. 2008, ACM: Singapore, Singapore. p. 459-466.
- [7] Pavkovic, M. and J. Protic. *The use of an intelligent forum crawler for data retrieval from e-learning portals*. in 6th International Conference on Education and New Learning Technologies (EDULEARN). 2014. Barcelona, SPAIN: Iated-Int Assoc Technology Education & Development.
- [8] Liu, W., H. Yan, and J. Xiao, *Automatically extracting user reviews from forum sites*. Comput. Math. Appl., 2011. 62(7): p. 2779-2792.
- [9] Song, X., et al., *Automatic extraction of web data records containing user-generated content*, in Proceedings of the 19th ACM international conference on Information and knowledge management. 2010, ACM: Toronto, ON, Canada. p. 39-48.
- [10] *Data Mining: Foundations and Intelligent Paradigms*. 1 ed. Intelligent Systems Reference Library, ed. D.E. Holmes and L.C. Jain. 2012: Springer-Verlag Berlin Heidelberg. 223-263.
- [11] Brewington, B.E. and G. Cybenko, *How dynamic is the Web?* Comput. Netw., 2000. 33(1-6): p. 257-276.

3. Полазне хипотезе

Полазне хипотезе које ће бити усвојене у овој дисертацији су:

- Развој модула и алгоритама за прикупљање искључиво новог садржаја приликом инкременталне посете дискусионој платформи је од великог значаја за уштеду ресурса окружења у којем се претраживач извршава.
- Детектовањем локације садржаја могуће је доћи до најновијег садржаја на ефикаснији начин. Распоред садржаја на дискусионој платформи је организован у зависности од тога како су странице сортиране. Детекцијом начина сортирања могуће је закључити на којим Веб странама ће најновији садржај бити генерисан.
- Коришћењем навигационих опција код дате технологије, односно постојећих скелетних линкова у одређеном редоследу, могуће је ефикасније циљати стране са најновијим садржајем.
- Различите технологије обезбеђују приказ различитих навигационих линкова, па самим тим и другачији пут до новог садржаја. Генерализовањем и минимизацијом ове путање се добија метод који на ефикасан начин покрива велики број дискусионих платформи.
- Модули и алгоритми морају да буду робусни тако да подржавају све врсте форум и Q&A технологија, како генеричке и јавно доступне, тако и специјализоване, ексклузивно примењене на једном Веб сајту.
- Детаљном анализом и евалуацијом постојећих Веб претраживача, што генеричких што специјализованих утврђује се да не постоје идеална решења која могу приликом инкременталног претраживања форума преузети искључиво нови садржај.

4. Научне методе истраживања

У овој дисертацији биће примењене следеће методе истраживања:

- Упоредна анализа постојећих специјализованих претраживача форума и систематизација њихових предности и недостатака у инкременталним условима претраживања. Евалуација специфичности и постојећих опција Веб сајт дискусионих технологија, како генерализованих тако и специјализованих.

- Софтверска имплементација којом ће се модули и алгоритми представљени у овој дисертацији имплементирати у виду целовитог софтверског система за инкрементално претраживање Веб форума. При томе ће, по потреби, бити коришћена постојећа решења за претрагу Веб форума и аутоматску екстракцију података. Приликом реализације ће се тежити објектно-оријентисаној парадигми.
- Симулациона евалуација, којом ће се проверити ефикасност постојећих алгоритама и модула кроз низ експеримента, у којима се мења тип технологије Веб форума, опциони навигациони линкови и начин презентације података.

5. Очекивани научни допринос

Очекивани научни допринос предложене докторске дисертације је у домену оптимизације свих аспеката инкременталног претраживања кориснички генерисаног садржаја на дискусионим платформама и имплементације добијених решења на већ постојеће претраживаче. Очекује се да ће ова дисертација довести до следећих научних доприноса:

- Биће урађен детаљан преглед и класификација већ постојећих специјализованих Веб форум претраживача.
- Извршиће се оптимизације у домену инкременталног претраживања: (1) боља искоришћеност пропусног опсега (2) смањено време извршавања (3) редукција дуплираног садржаја (4) боља искоришћеност ресурса система на којем се претраживач извршава. Коришћењем метода и алгоритама добијених овим истраживањем, где се циљано прескаче већ претходно посећени садржај, оптимизоваће се и процес претраживања.
- Постиглићи ће се боља искоришћености постојећих навигационих путања и опција које технологија која презентује кориснички генерисан садржај поседује.
- Биће развијен и имплементиран комплетни инкрементални претраживач Веб форума, базиран на FoCUS специјализованом претраживачу и WeRE аутоматској екстракцији кориснички генерисаних података. Извршиће се проширење и унапређење система, тако да се интелигентно адаптира навигационим и опционим линковима, начину презентовања и сортирања садржаја на Веб форум странама. Имплементираће се алгоритам за динамичко одређивање минималног пута до најновијег садржаја.
- Спровешће се испитивање скалабилности и евалуација перформанси добијеног решења симулационом техником генерисања садржаја Веб форума у различитим временским периодима, на великом броју различитих Веб форум технологија и сајтова.

6. План истраживања

Планира се да се истраживање изведе у следећим фазама:

- Прва фаза обухватиће преглед и евалуацију специфичности постојећих дискусионих технологија, како опште наменских тако и специјализованих. Планирана је упоредна анализа и класификација презентације садржаја, постојећих навигационих опција и начина сортирања.
- Друга фаза обухватиће преглед и класификацију постојећих специјализованих Веб форум претраживача као и решења за аутоматску екстракцију кориснички генерисаног садржаја који ће касније бити коришћени у изради система за инкрементално претраживање. У обзир ће бити узета само она решења која на што ефикаснији начин прате скелетне линкове и заобилазе неважне стране Веб сајта.

- Трећа фаза предвиђа имплементацију основе система која се заснива на изабраним већ постојећим решењима за претрагу и екстракцију података.
- Четврта фаза предвиђа проширење и надоградњу основног система модулима и алгоритмима за прикупљање нових података без приступања странама на којима је искључиво садржај прикупљен у претходним посетама.
- Пета фаза обухватиће симулациону анализу предложеног система и његових модула, као и оцену и дискусију добијених резултата.

На основу резултата наведених пет фаза истраживања биће постављене основе за квалитетније и ефикасније прикупљање кориснички генерисаног садржаја приликом инкременталног претраживања са посебним акцентом на детекцији презентације садржаја, чиме ће се проширити могућности за будућа унапређења већ постојећих претраживача.

7. Закључак и предлог

На основу свега изложеног, Комисија констатује да кандидат Милош Павковић испуњава потребне услове да може приступити изради докторске дисертације, под радним насловом **“Побољшање перформанси прикупљања кориснички генерисаних садржаја на Вебу применом адаптивних интелигентних метода”**.

Ужа научна област теме је Софтверско инжењерство, за које је Електротехнички факултет у Београду матичан. За ментора се предлаже проф. др Јелица Протић, редовни професор, запослена на Електротехничком факултету Универзитета у Београду. Уз овај Извештај се прилаже списак радова који проф. др Јелицу Протић квалификују за ментора.

Узимајући у обзир све наведене чињенице, Комисија предлаже Наставно-научном већу Електротехничког факултета у Београду да одобри кандидату Милошу Павковићу израду докторске дисертације под горе наведеним насловом, под руководством ментора проф. др Јелице Протић.

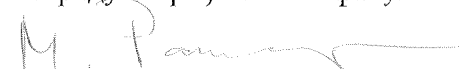
У Београду, 27.12.2017. године

ЧЛАНОВИ КОМИСИЈЕ


др Јелица Протић, редовни професор
Универзитет у Београду – Електротехнички факултет


др Бошко Николић, редовни професор
Универзитет у Београду – Електротехнички факултет


др Милош Ковачевић, ванредни професор
Универзитет у Београду – Грађевински факултет


др Марија Рашајски, ванредни професор
Универзитет у Београду – Електротехнички факултет