

**НАСТАВНО-НАУЧНОМВЕЋУ
ГРАЂЕВИНСКОГ ФАКУЛТЕТА
УНИВЕРЗИТЕТА У БЕОГРАДУ**

Одлуком Наставно-научног већа Грађевинског факултета Универзитета у Београду бр. 154/10-16 од 11.04.2018. године именовани смо за чланове Комисије за преглед, оцену и одбрану докторске дисертације кандидата Ђорђа Недељковића, дипл.грађ.инж., под насловом:

**ИЗДВАЈАЊЕ И ВИЗУЕЛИЗАЦИЈА ЗНАЊА ИЗ ТЕКСТУАЛНИХ ИЗВОРА ЗА ПОТРЕБЕ
УПРАВЉАЊА ИНВЕСТИЦИОНИМ ПРОЈЕКТИМА У ГРАЂЕВИНАРСТВУ**

После прегледа достављене дисертације Комисија подноси Научно–наставном већу Грађевинског факултета следећи

ИЗВЕШТАЈ

1. УВОД

1.1. Хронологија одобравања и израде дисертације

На седници Катедре за управљање пројектима у грађевинарству одржаној 07.04.2016. Ђорђе Недељковић је јавно излагао предложену тему докторске дисертације под насловом „Издвајање и визуелизација знања из текстуалних извора за потребе управљања инвестиционим пројектима у грађевинарству“. Комисија у саставу в.проф. др Милош Ковачевић, проф. др Бранислав Ивковић, проф. др Вељко Милутиновић (са Електротехничког Факултета Универзитета у Београду), и в.проф. др Ненад Иванишевић је прихватила предложену тему.

Одлуком Наставно-научног већа Грађевинског факултета бр. 154/3 од 22.04.2016. године, одређена је Комисија за оцену научне заснованости теме докторске дисертације под насловом „Издвајање и визуелизација знања из текстуалних извора за потребе управљања инвестиционим пројектима у грађевинарству “ у саставу в.проф. др Милош Ковачевић, проф. др Бранислав Ивковић, проф. др Вељко Милутиновић (са Електротехничког Факултета Универзитета у Београду), и в.проф. др Ненад Иванишевић. Позитиван извештај Комисије за оцену научне заснованости теме докторске дисертације усвојен је на седници Наставно-научног већа Грађевинског факултета одржаној 19.05.2016. године (одлука бр. 154/5 од 20.05.2016. године).

Веће научних области грађевинско-урбанистичких наука Универзитета у Београду на седници одржаној 28.06.2016. (одлука бр. 61206-2610/2-16 од 28.06.2016. године) усвојило је

предлог теме докторске дисертације кандидата Ђорђа Недељковића.

Кандидат је предао завршену докторску дисертацију Служби за студентска питања Грађевинског факултета 30.03.2018. године.

Наставно-научно веће Грађевинског факултета на седници одржаној 04.05.2018. године именовало је Комисију (одлука бр. 154/10-16 од 11.05.2018.) за оцену и одбрану докторске дисертације у следећем саставу:

1. др Милош Ковачевић, ванредни професор
Универзитет у Београду, Грађевински факултет
2. др Ненад Иванишевић, ванредни професор
Универзитет у Београду, Грађевински факултет
3. др Наташа Прашчевић, ванредни професор
Универзитет у Београду, Грађевински факултет
4. др Бранислав Ивковић, редовни професор у пензији
Универзитет у Београду, Грађевински факултет
5. др Вељко Милутиновић, редовни професор у пензији
Универзитет у Београду, Електротехнички факултет

1.2. Научна област дисертације

Тема докторске дисертације припада научној области *Грађевинарство* и ужој научној области *Примена информационих технологија у грађевинарству и геодезији*, која је дефинисана Статутом Грађевинског факултета Универзитета у Београду. Ментор дисертације је др Милош Ковачевић, ванредни професор Грађевинског факултета Универзитета у Београду.

1.3. Биографија кандидата

Ђорђе Недељковић рођен је 12.11.1984. године у Сремској Митровици. У Смедеревској Паланци је завршио основну школу и гимназију природно-математичког смера. Грађевински факултет Универзитета у Београду уписао је школске 2003/04. године, а дипломирао је 2009. године на конструктивном смеру, са просечном оценом 8,61 и оценом 10 на дипломском раду под називом „Процена тржишне вредности грађевинских објеката применом техника машинског учења“. Докторске студије је уписао крајем 2009. год. Све предвиђене испите положио је закључно са октобром 2014. године, са просечном оценом 9.625. Ради као асистент-студент докторских студија на предметима Рачунарско цртање у грађевинарству, Основе програмирања у VisualBasic-у, Основе програмирања у Пајтону и Објектно оријентисано програмирање.

Као аутор и коаутор, до сада је публикувао 9 радова у међународним часописима и зборницима међународних конференција, од чега један у часопису са СЦИ листе (категорије M21). Учествовао је као истраживач на технолошким пројектима „Примена GNSS и LIDAR технологије у мониторингу стабилности инфраструктурних објеката и терена“ и „Истраживање стања и метода унапређења грађевинских конструкција са аспекта употребљивости, носивости, економичности и одржавања“. Коаутор је књиге „Основе програмирања у Matlab-у, збирка задатака“ (издање 2010. год.). Говори, чита и пише енглески језик.

2. ОПИС ДИСЕРТАЦИЈЕ

2.1. Садржај дисертације

Докторска дисертација Ђорђа Недељковића, под насловом „Издвајање и визуелизација знања из текстуалних извора за потребе управљања инвестиционим пројектима у грађевинарству“, има 149 страна, садржи 40 слика и 17 табела. На почетку дисертације дата је изјава захвалности, резиме са кључним речима и садржај. Дисертација је подељена на 10 глава:

1. Уводна разматрања
2. Коришћење неструктурираних текстуалних информација у грађевинском сектору Републике Србије
3. Постојећи системи за рад са документима у оквиру којих се изводе информације за потребе управљања пројектом
4. Предложени приступ за издвајање знања – претпоставке и ограничења
5. Аутоматска детекција значајних фраза из текстуалних извора
6. Погодне репрезентације знања
7. Предложена репрезентација информација
8. Складиштење и приступање репрезентацији значајних фраза
9. Примене графа значајних фраза у окружењу инвестиционог пројекта
10. Закључна разматрања

Списак цитиране литературе садржи 116 наслова. Биографија кандидата дата је на крају дисертације.

Дисертација је технички обликована према упутствима Сената Универзитета у Београду и посебним упутствима за обликовање штампане и електронске верзије доктората. Садржи обавезна поглавља и обрасце: изјава о ауторству, изјава о истоветности електронске и штампане верзије, и изјава о коришћењу.

2.2. Кратак приказ појединачних поглавља

Основни текст разматране докторске дисертације има десет глава.

У уводној **глави 1** су дефинисани основни појмови и задаци истраживања докторске дисертације.

Анализа тренутне праксе у домену коришћења информација из неструктурираног текста, у грађевинском сектору Републике Србије, приказана је у **глави 2**. Истраживање је спроведено кроз две анкете у којима су испитани постојећи поступци у раду са документима. Истражено је како грађевинска предузећа обављају интерну и екстерну комуникацију, као и како се обрађују подаци значајни за доношење одлука.

У **глави 3** размотрени су постојећи системи за рад са документима који настају током животног циклуса пројекта, са становишта могућности издвајања информација неопходних за управљање пројектом. Показане су специфичности поступка извођења новог знања у окружењу система датотека, система за управљање информацијама, као и система заснованог на онтологији.

На основу анализе постојећих решења за издвајање значајних информација из неструктурираног текста, у **глави 4** су формулисане основне претпоставке, описан концепт и наведена ограничења предложеног поступка за издвајање и визуелизацију знања из текстуалних извора.

У **глави 5** је дефинисан поступак издвајања релевантних концепата из неструктурираних докумената на пројекту. Концепти су представљени значајним паровима речи које се издвајају применом језички независних статистичких мера за одређивање међузависности. Предложен је поступак за филтрирање неинформативних парова, заснован на ентропији контекста појављивања. На овом месту, извршена је експериментална провера свих метода предложеног поступка, као и резултати експеримента којим се утврђује семантички капацитет парова да опишу препознате концепте на пројекту.

Глава 6 разматра постојеће репрезентације које би представљале основу за организовање издвојених концепата у погодну структуру за извођење нових знања. Разматране су семантичке мреже, концептуални графови и оквири.

Доменски и језички независан поступак конструисања репрезентације издвојених концепата, као графа значајних фраза, описан је у **глави 7**. Дефинисан је критеријум за успостављање релација између значајних парова, заснован на заједничком контексту појављивања унутар докумената у корпусу. Добијени граф користи се за проналажење значајних фраза састављених од већег броја речи, као и за додатно рангирање фраза према динамичности суседа.

У глави 8 се разматрају погодна окружења за складиштење и обраду предложене репрезентације. Поред се релациона и графовска база података, са становишта брзине извршавања и једноставности задавања упита. Посебно се разматра могућност визуелизације резултата упита, као неопходне опције која омогућава правилну корисничку интерпретацију приликом извођења нових знања.

Глава 9 описује интерактивни рад у окружењу графа значајних фраза. Дефинисани су поступци за одређивање блиских значајних концепата на пројекту, као и њихово праћење кроз време. Предложена је хеуристика за одређивање комплексних концепата који покривају одређене теме на пројекту. Описано је могуће семантичко проширивање предложене репрезентације, увођењем кориснички дефинисаних ентитета, попут датума, особе и акције. Проширивање се може обавити уз минимално претходно ангажовање експерта. Могућности проширене репрезентације илустроване су на примеру одређивања интеракције између учесника на пројекту, а на основу обраде записника са састанака. На овом месту, размотрени су потенцијални недостаци предложеног приступа који проистичу из потребе да експерт, вођен когнитивном пристрасношћу, погрешно интерпретира добијене резултате.

Дисертацију закључује глава 10, у којој су дата закључна разматрања о предложеном решењу и наведени предлози могућих праваца за будуће истраживање.

3. ОЦЕНА ДИСЕРТАЦИЈЕ

3.1. Савременост и оригиналност

За успешно управљање инвестиционим пројектом неопходно је ефикасно праћење и контролисање протока информација. Да би у сваком тренутку успешно управљали пројектом, експерти морају имати тачне, јасне и правовремене информације.

Највећи део садржаја који се користи за доношење одлука на инвестиционом пројекту је у форми неструктурираних података. Неструктурирани подаци који се најчешће налазе у текстуалном облику, чине око 80% пословних информација у предузећу. Ови подаци налазе се у форми која није погодна за извођење нових знања (одсуство формалног описа и структуре), већ су корисници на пројекту принуђени да уз помоћ класичних алата претраге комбинују информације да би формирали ширу слику.

Како би се савладао проблем презасићености информацијама на пројекту, неопходно је коришћење одговарајућих информационих система за рад са документима. Међутим, највећи део информационих система који се користе у грађевинској индустрији, заснива се на складиштењу које по природи одговара класичним папирним формама. Када је у питању издавање знања из неструктурираних текстуалних података, постојећи системи не могу адекватно да одговоре на све специфичности окружења грађевинског пројекта.

У дисертацији је описан приступ за издвајање релевантних концепата на пројекту у виду графа значајних фраза. Ова структура захтева интерпретацију од стране крајњег корисника, али је однос између уложеног труда за успостављање система и труда при закључивању у току коришћења, повољнији у односу на постојеће системе за рад са документима. У истраживању је претпостављено да се знање на пројекту може представити као скуп детектованих концепата из различитих категорија, повезаних предефинисаним релацијама, насталих из више текстуалних извора током времена. Предложен је аутоматски поступак за издвајање значајних концепата из неструктурираних и полуструктурираних текстуалних извора. Поред тога, предложен је и начин њиховог организовања у репрезентацију погодну за визуелизацију и извођење знања интерпретацијом од стране експерта.

Суштинска особина предложеног процеса за аутоматско издвајања релевантних концепата из неструктурираних текстуалних извора је да не зависи од претходно дефинисаног експертског знања (или да зависност буде минимална). Предложени приступ је трансферабилан јер не захтева значајне ресурсе и прилагођавања за различите пројекте. У потпуности је прилагодљив постојећим пословним процесима корисника у раду са текстуалним изворима. Приступ је независан у односу на језик документа јер су поступци за детекцију основних концепата-фраза засновани на статистичким методама, које су универзално применљиве у сваком језичком окружењу.

Издвојена репрезентација информација је имплементирана у графовској бази података, што корисницима омогућава да детектују и визуелизују различите скривене обрасце у подацима. Помоћу визуелизације издвојених образаца, учесници могу да сагледају текуће трендове на пројекту, што их додатно мотивише да истражују знање похрањено у неструктурираним подацима.

3.2. Осврт на референтну и коришћену литературу

У изради ове докторске дисертације је коришћено 116 библиографских јединица. Већину референци чине радови објављени у водећим међународним часописима попут *Journal of Computing in Civil Engineering*, *Automation in Construction*, *Journal of Construction Engineering and Management*, књиге референтних аутора из ове области, као и радови објављени на значајним међународним конференцијама. Највећи број референци је новијег датума: 97 референци је публиковано након 2000. године, од чега 54 између 2010. и 2017. године.

3.3.Опис и адекватност примењених научних метода

У дисертацији су коришћене опште научне методе из области претраживања информација (Information retrieval), интелигентне анализе текста (Text mining), обраде природног језика (Natural Language Processing) и теорије графова (Graph Theory).

Да би се утврдило како се користе информације из неструктурираних текстуалних извора у грађевинском сектору Републике Србије, примењена је емпиријска метода квантитативног истраживања (анкетни упитници). Приликом обраде резултата анкета коришћене су адекватне статистичке методе.

За детекцију и рангирање значајних фраза из текстуалних докумената, коришћене су статистичке мере за одређивање семантичке повезаности речи.

У анализи експерименталних резултата за детекцију и рангирање значајних фраза, коришћене су методе за евалуацију модела (одзив, прецизност, криве оперативне карактеристике пријемника) које се примењују у области претраживања информација.

Методе теорије информација и методе динамичке анализе графова примењене су за одређивање информативности текстуалног контекста. Приликом конструисања значајних фраза које садрже више од две речи, коришћен је Bron-Kerbosch алгоритам за одређивање максималних клика у графу.

Значајност информација садржаних у дворечним фразама из докумената на пројекту потврђена је техникама груписања документа (хијерархијско груписање коришћењем графова).

Наведене методе истраживања су актуелне и у потпуности примерене за примену у предметном истраживању.

3.4. Применљивост остварених резултата

Предложена репрезентације конструисана је из података са два велика инвестициона пројекта. Приказан је поступак за одређивање повезаних концепата на пројекту, као и њихово праћење кроз време. Наведени поступак омогућава да се идентификују теме од интереса и трендови груписања информација на пројекту, што може помоћи у идентификацији ситуација које треба додатно истражити. Предложена је хеуристика за одређивање комплексних концепата који покривају одређене теме и који се могу искористити за аутоматско обележавање докумената према њиховој заступљености. Описано је могуће семантичко проширивање предложене репрезентације, увођењем кориснички дефинисаних ентитета попут датума, особе и акције. Могућности проширене репрезентације илустроване су на примеру одређивања интеракције између појединих типова учесника на пројекту (обим и структура комуникације).

3.5. Оцена достигнутих способности кандидата за самостални научни рад

Кандидат се у оквиру своје докторске дисертације бавио изучавањем и критичком анализом доступне референтне литературе, планирањем, спровођењем, анализом и верификацијом резултата експерименталног истраживања, као и моделирањем предложеног решења. Систематичним приступом постављеном проблему, повезујући различите сегменте научно-истраживачког рада, Ђорђе Недељковић је успешно решио постављене задатке и доказао да поседује способност за самостални научно-истраживачки рад.

4. ОСТВАРЕНИ НАУЧНИ ДОПРИНОС

4.1. Приказ остварених научних доприноса

У оквиру докторске дисертације Ђорђа Недељковића остварени су следећи научни доприноси:

1. Извршено је поређење статистичких метода за аутоматску детекцију и рангирање значајних дворечних фраза из корпуса неструктурираних текстуалних извора на пројекту;
2. Имајући у виду различите преференције појединих статистичких метода, предложен је поступак за побољшану детекцију и рангирање значајних дворечних фраза, заснован на комбиновању полазних метода;
3. Дефинисана је нова метода детекције значајних дворечних фраза која користи меру ентропије скупа суседстава фразе. Ентропија представља корективни фактор у појединим статистичким методама, који уклања неинформативне фразе чије су речи статистички корелисане;
4. Доказана је значајност информација садржаних у дворечним фразама, када су у питању документи са инвестиционих пројеката;
5. Дефинисана је графовска репрезентација дворечних фраза, заснована на релацији припадности сличним контекстима. Репрезентација омогућава издвајање вишеречних фраза применом Bron-Kerbosch алгоритма за детекцију максималних клика у графу;
6. Предложена је употреба мере динамичности суседа у графу фраза, који еволуира кроз време, за додатно рангирање и филтрирање информативних фраза;
7. Дефинисана је хеуристика за итеративну конструкцију графа комплексног концепта, којом се проналазе комплексни концепти који покривају одређене теме на пројекту.

4.2. Критичка анализа резултата истраживања

Применљивост решења на различите домене проблема (пројекте) је обезбеђена тиме што се издвојени релевантни концепти аутоматски организују у структуру која не захтева претходну конфигурацију од стране експерта. Иако би постојање онтологије обезбедило потенцијално богатију репрезентацију информација, као и формалну логику за закључивање, такав приступ би био ограничен само на домен проблема покривен одговарајућом онтологијом. Предложена репрезентација се у потпуности ослања на експертску интерпретацију којом се, из структурираних релевантних концепата и веза између њих, изводи ново знање. Међутим, како у предложеном приступу није дефинисан механизам којим се испитује валидност изведених закључака, могуће су грешке у интерпретацији услед когнитивне пристрасности. Овај проблем може се делимично превазићи применом различитих алата за бољи увид у семантичке контексте информације (повезаност са изворним документима), одговарајућу визуелизацију и коришћење више аналитичких приступа у паралели, на шта је кандидат и указао у тексту дисертације.

Велики инвестициони пројекти обично генеришу вишејезичне документе. Аналогно применљивости на различитим пројектима, функционисање у сваком језичком окружењу је обезбеђено тиме што се основни поступак, којим се издвајају значајне фразе, не заснива на методама обраде природног језика. Фокус на неколико познатих језика, који су на одговарајући начин покривени ресурсима за напредну обраду текста, дао би боље резултате од метода које су засноване само на статистичким мерама корелисаности речи. Све мере за детекцију значајних фраза су показале побољшање перформанси када се укључе технике за обраду природног језика. Предложени приступ узима у обзир ову чињеницу и дефинисан је тако да може да укључи одговарајуће језичке ресурсе. У истраживању је експериментално утврђено да је разлика између најбољих метода које користе и не користе алате за обраду природног језика у прихватљивим границама. Ова особина предложеног решења је посебно значајна за српско грађевинарство, где не постоје одговарајући језички ресурси.

4.3. Верификација научних доприноса

У току израде дисертације, Ђорђе Недељковић је међународној и домаћој, научној и стручној јавности представио свој рад кроз следеће публикације:

Категорија M21:

1. Ђорђе Nedeljković, Miloš Kovačević (2017) Building a Construction Project Key-Phrase Network from Unstructured Text Documents. *Journal of Computing in Civil Engineering*. 31 (6)

Категорија M33:

1. Ђорђе Nedeljković, Miloš Kovačević (2015) Detecting Concepts in Construction Project Documents using Statistical Measures for Semantic Similarity. In: *Proceedings of the Fifteenth International Conference on Civil, Structural and Environmental Engineering Computing*.

Категорија M34:

1. Ђорђе Nedeljković, Miloš Kovačević (2015) Presek stanja u građevinskom sektoru Srbije po pitanju korišćenja alata za upravljanje dokumentima. In: *Project Society Conference 2016*.

5. ЗАКЉУЧАК И ПРЕДЛОГ

На основу анализе докторске дисертације, примењених научних метода, остварених резултата истраживања и научних доприноса, може се констатовати да докторска дисертација Ђорђа Недељковића, дипл.грађ.инж, под насловом „Издавање и визуелизација знања из текстуалних извора за потребе управљања инвестиционим пројектима у грађевинарству“, представља оригинални научни рад, који има значајан научни допринос у области грађевинарства. У дисертацији се дефинишу поступци за аутоматско издавање и репрезентовање значајних информација из неструктурираних текстуалних извора са инвестиционог пројекта. Предложено решење је применљиво на различите пројекте и у различитим језичким окружењима, уз минимални труд експерта за конфигурацију система. Решење омогућава да се уз помоћ визуелизације анализирају везе између значајних концепата издвојених из неструктурираних извора. Предложени поступци верификовани су путем експеримената у складу са опште прихваћеном научном методологијом.

Комисија сматра да урађена докторска дисертација Ђорђа Недељковића, дипл.грађ.инж у потпуности испуњава све захтеване критеријуме и да је кандидат показао способност за самостално бављење научно-истраживачким радом

Имајући у виду све што је напред речено, Комисија сматра да урађена докторска дисертација Ђорђа Недељковића, дипл.грађ.инж. представља оригиналан и вредан научни допринос у области примене информационих технологија у грађевинарству и предлаже Наставно-научном већу Грађевинског факултета Универзитета у Београду да прихвати позитивну оцену докторске дисертације кандидата Ђорђа Недељковића под насловом „Издвајање и визуелизација знања из текстуалних извора за потребе управљања инвестиционим пројектима у грађевинарству“ и да упути захтев Већу научних области грађевинско-урбанистичких наука Универзитета у Београду за давање сагласности за јавну одбрану дисертације.

Београд, 27.04.2018.

ЧЛАНОВИ КОМИСИЈЕ

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

.....
др Ненад Иванишевић, ванредни професор
Универзитет у Београду, Грађевински факултет

.....
др Наташа Прашчевић, ванредни професор
Универзитет у Београду, Грађевински факултет

.....
др Бранислав Ивковић, редовни професор у пензији
Универзитет у Београду, Грађевински факултет

.....
др Вељко Милутиновић, редовни професор у пензији
Универзитет у Београду, Електротехнички факултет