

**УНИВЕРЗИТЕТ У НОВОМ САДУ**  
**ТЕХНИЧКИ ФАКУЛТЕТ "МИХАЈЛО ПУПИН"**  
**ЗРЕЊАНИН**

**ОБРАЗАЦ 6.**

**ИЗВЕШТАЈ О ОЦЕНИ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ**

**-обавезна садржина- свака рубрика мора бити попуњена**

(сви подаци уписују се у одговарајућу рубрику, а назив и место рубрике не могу се мењати или изоставити)

<b>I ПОДАЦИ О КОМИСИЈИ</b>
1. Датум и орган који је именовано комисију Дана <b>18. 05. 2016.</b> године Наставно-научно веће Техничког факултета „Михајло Пупин“ у Зрењанину је именовало Комисију за оцену и одбрану докторске дисертације
2. Састав комисије са назнаком имена и презимена сваког члана, звања, назива уже научне области за коју је изабран у звање, датума избора у звање и назив факултета, установе у којој је члан комисије запослен: 1. Проф. др Миодраг Ивковић – редовни професор, Информационе технологије, 29.12.2011. године, Технички факултет “Михајло Пупин”, Зрењанин, председник 2. Проф. др Драгана Глушац, редовни професор, Информационе технологије у образовању, 18.11.2015. године, Технички факултет “Михајло Пупин”, Зрењанин, члан 3. Доц. др Срђан Попов, доцент, ФТН, Нови Сад, Примењене рачунарске науке и информатика, 26.4.2012.. године, Факултет техничких наука Нови Сад, члан 4. Доц. др Здравко Иванковић, доцент, Информационе технологије, 1.9.2015.. године, Технички факултет “Михајло Пупин”, Зрењанин, члан 5. Проф. др Бранко Маркоски, ванредни професор, Информационе технологије, 10.02.2014. године, Технички факултет “Михајло Пупин”, Зрењанин, ментор
<b>II ПОДАЦИ О КАНДИДАТУ</b>
1. Име, име једног родитеља, презиме: Војкан, Радмило, Николић
2. Датум рођења, општина, држава: 29.05.1971. Алексинац, Алексинац, Србија
3. Назив факултета, назив студијског програма дипломских академских студија – мастер и стечени стручни назив Електронски факултет, Универзитет у Нишу, Рачунарска техника и информатика, дипломирани инжењер електротехнике на профилу за рачунарску технику и информатику
4. Година уписа на докторске студије и назив студијског програма докторских студија
5. Назив факултета, назив магистарске тезе, научна област и датум одбране: 07.10.2010., Технички факултет, Бор, Индустијска информатика, „Развој информационих система за мала и средња предузећа“
6. Научна област из које је стечено академско звање магистра наука: Индустијска информатика
<b>III НАСЛОВ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:</b> „МОДЕЛОВАЊЕ И ПРЕТРАЖИВАЊЕ НАД НЕСТРУКТУИРАНИМ ПОДАЦИМА И ДОКУМЕНТИМА У Е-УПРАВИ РЕПУБЛИКЕ СРБИЈЕ“
<b>IV ПРЕГЛЕД ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:</b> Навести кратак садржај са назнаком броја страна, поглавља, слика, шема, графикона и сл. Докторска дисертација састоји се од седам поглавља, листе слика, листе табела и списка референци,

укључујући 44 табеле и 63 слике, 130 референци на укупно 156 страна. Предмет истраживања докторске дисертације је одабир најпогодније технологије за израду модела за брзо проналажење корисних информација и вредновање система. Основна поставка истраживања је да је могуће креирати модел за претраживање неструктурираних података и докумената у е-Управи Републике Србије. Поред тога, помоћне премисе су: могуће је груписати неструктуриране документе на основу тражења података у документима (*metadata*) и могуће је реализовати *question answering* систем базиран на ВоС моделу који даје брзе одговоре на упите грађана у вези чланова Кривичног законика Републике Србије. Циљ истраживања је проналажење ефикаснијег начина за организовање садржаја недовољно развијене структуре текста који се налазе као вредности поља табела у базама података или као документи у фајл систему. За реализацију истраживања креиран је *question answering* систем базиран на ВоС моделу и алгоритам дела система за класификацију упита за чланове кривичних дела Кривичног законика. Садржај дисертације обухвата:

Листа слика

Листа табела

Коришћене ознаке

## **1. Увод**

- 1.1. Потребе истраживања
- 1.2. Проблем истраживања
- 1.3. Предмет и циљ истраживања
- 1.4. Задаци истраживања
- 1.5. Хипотезе истраживања
- 1.6. Програм истраживања
- 1.7. Методолошки концепт
  - 1.7.1. Начин избора, величина и конструкција узорка
  - 1.7.2. Технике, поступци и мерни инструменти истраживања
  - 1.7.3. Место експерименталног истраживања
- 1.8. Структура дисертације

## **2. е-Управа Републике Србије**

- 2.1. Интероперабилност е-Управе у контексту Европског оквира интероперабилности
  - 2.1.1. Интероперабилност у е-Управи
  - 2.1.2. Европски оквир интероперабилности
  - 2.1.3. Концептуални модел EIF
  - 2.1.4. Нивои интероперабилности у EIF
  - 2.1.5. Реализација интероперабилности помоћу GSB у е-Управи Републике Србије
- 2.2. Национални оквир интероперабилности Републике Србије и Сервисно оријентисана архитектура
  - 2.2.1. Национални оквир интероперабилности Републике Србије
  - 2.2.2. СОА на бази веб-сервиса и интероперабилност
  - 2.2.3. Концепт е-Управе Републике Србије
- 2.3. Г2Г интеграција МУП-а Републике Србије са порталом е-Управа
  - 2.3.1. Екстранет МУП-а Републике Србије
  - 2.3.2. Случајеви коришћења Екстранета МУП-а Републике Србије
- 2.4. Безбедност е-Управе Републике Србије
  - 2.4.1. „Листа стандарда интероперабилности“ е-Управе Републике Србије
  - 2.4.2. Безбедност е-Управе Републике Србије

## **3. Question answering системи**

- 3.1. ТМ и е-Управа

3.1.1.	ТМ апликације у е-Управи
3.1.2.	Дубинска анализа података
3.1.3.	Технике ТМ
3.2.	Анализа максималне учесталости секвенци речи (eng. Mining Maximal Frequent Word Sequences)
3.3.	Рангирање резултата претраге (eng. ranking score)
3.4.	Анализа података садржаних у вишејезичним текстовима (eng. Multilingual Text Mining - MLTM)
3.5.	Софтверски пакети за ТМ
3.6.	Постојећи оквири за QA системе
<b>4.</b>	<b>Apache Lucene</b>
4.1	Архитектура Apache Lucene
4.1.1	Концепт Apache Lucene
4.1.2	Преглед композиционе структуре Apache Lucene
4.2	Lucene документ
4.2.1	Package org.apache.lucene.document
4.3	Lucene анализа
4.3.1	Package org.apache.lucene.analysis
4.3.2	Аналајзер
4.3.3	Токенизер
4.4	Lucene индексирање
4.5	Lucene претраживање
4.5.1	Package org.apache.lucene.QueryParser
4.5.2	Package org.apache.lucene.Search
4.5.3	Алгоритам Lucene за претраживање индекса
4.6	LUKE
<b>5.</b>	<b>Моделовање система за добијање брзих одговора за сервисе е-Управе Републике Србије у области Кривичног законика</b>
5.1.	Принцип коришћења система е-Управа
5.1.1.	Теоријске основе за развој система брзих одговора
5.1.2.	Приказ стања у Кривичном законика Републике Србије
5.1.3.	Концепти поређења кратког текста: поређење чланова Кривичног законика
5.1.4.	Принцип претраживања применом MathML стандарда
5.2.	Предложени оквир за реализацију могућег QA веб-сервиса
5.2.1.	Стоп-речи
5.2.2.	QA систем базиран на ВоС моделу
<b>6.</b>	<b>Анализа експерименталних резултата</b>
6.1.	Приказ текстуалних докумената у векторском облику
6.2.	Провера прецизности предложеног система
<b>7.</b>	<b>Закључак</b>
	Литература
<b>V</b>	<b>ВРЕДНОВАЊЕ ПОЈЕДИНИХ ДЕЛОВА ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:</b>
	Докторска дисертација се састоји од 7 поглавља. У уводном поглављу на 9 страна указано је на главне проблеме који су присутни у процесу претраживања неструктурираних података, описан је предмет истраживања и постављен циљ

дисертације.

Друго поглавље (32 стране) даје преглед концепата е-Управе уопште, као и концепт и архитектуру е-Управе у Републици Србији. Представљени су концепти Европског и Националног оквира интероперабилности, Сервисно оријентисана архитектура, ЕСБ и ГСБ и њихова имплементација у е-Управи Републике Србије. У овом поглављу су приказана и конкретна решења и сервиси који су имплементирани у е-Управи Републике Србије. Један део овог поглавља се односи на безбедност е-Управе Републике Србије, као и на безбедност електронских сервиса, која се базира на примени „Листе стандарда интероперабилности“.

Треће поглавље на 15 страна даје теоријске основе дубинске анализе текста и *question answering* системима и прегледом литературе ових области. Представљени су и *question answering* системи, оквири за развој *question answering* система и софтверски пакети за дубинску анализу текста. Поред тога, дат је преглед апликација е-Управа Влада разних земаља које користе дубинску анализу текста.

Четврто поглавље на 30 страна даје преглед концепта и архитектуре софтверске библиотеке *Apache Lucene*, која је коришћена у овој докторској дисертацији како би се обезбедио процес индексирања и претраживања докумената. Ова библиотека се састоји од веома комплексних функција, чији су алгоритми и механизми веома сложени. Овде су представљене основне функције *Apache Lucene* и начини њиховог коришћења у апликацијама. На крају је представљен LUKE графички едитор за преглед и анализу индекса које је претходно креирао *Apache Lucene*.

У петом поглављу „Моделовање система за добијање брзих одговора за сервисе е-Управе Републике Србије у области Кривичног законика“ које има 27 страна, говори се о концепту *question answering* система базираном на ВоС моделу и алгоритаму за класификацију упита за чланове кривичних дела Кривичног законика Републике Србије. Детаљно је представљен и описан *question answering* систем са свим својим компонентама и могућностима, као и алгоритам дела система за класификацију одговора у оквиру чланова за кривична дела из Кривичног законика.

Шесто поглавље које укупно има 18 страна, приказује резултате истраживања и даје анализу добијених резултата.

Седмо, завршно поглавље, сумира резултате рада, даје закључна разматрања и указује на могуће даље правце истраживања. На крају су дате референце са списком од укупно 131 литературног извора.

## **VI СПИСАК НАУЧНИХ И СТРУЧНИХ РАДОВА КОЈИ СУ ОБЈАВЉЕНИ ИЛИ ПРИХВАЋЕНИ ЗА ОБЈАВЉИВАЊЕ НА ОСНОВУ РЕЗУЛТАТА ИСТРАЖИВАЊА У ОКВИРУ РАДА НА ДОКТОРСКОЈ ДИСЕРТАЦИЈИ**

Таксативно навести називе радова, где и када су објављени. Прво навести најмање један рад објављен или прихваћен за објављивање у часопису са ISI листе односно са листе министарства надлежног за науку када су у питању друштвено-хуманистичке науке или радове који могу заменити овај услов до 01. јануара 2012. године. У случају радова прихваћених за објављивање, таксативно навести називе радова, где и када ће бити објављени и приложити потврду о томе.

Од 56 радова кандидата мр Војкана Николића, 17 се односи директно на истраживања у оквиру рада на докторској дисертацији. Кандидат има 1 рад објављен у часопису са импакт фактором а 16 радова је презентовано на међународним и домаћим конференцијама.

### **M23 – Рад у часопису међународног значаја**

1. Z. Stević, M. Rajčić-Vujanović, I. Radovanović, V. Nikolić, Modeling and Sensing of Electrochemical Processes upon Dirac Potentiostatic Excitation of Capacitive Charging/Discharging, Int. J. Electrochem. Sci., 10 (2015) 6020-6029; ISSN:1452-3981; IF(2014) = 1.500; M23 (19/28).

Кандидат је први аутор у 13 референци које се односе на истраживања у дисертацији.

### **M33 – Рад саопштен на скупу међународног значаја штампан у целини**

1. V. Nikolić, M. Ivković, P. Djikanović, S. Nedeljković, Realizacija procesa analize tekstualnih dokumenata pomoću biblioteke otvorenog koda Apache Lucene, DQM 2016, Prijedor, Republika Srbija
2. V. Nikolić, B. Markoski, K. Kuk, D. Randjelović, M. Ivković, Possibilities of application of

intelligent search techniques in e-Government services of the Republic of Serbia, AIIT 2016, Bitola, FYROM

3. **V. Nikolić**, P. Djikanović, S. Nedeljković, Tehniques Of Cyberspace Information Searching In Serbian Text Document: Case Study For Crime Law, Archibald Rais, KPA Beograd 2016, Republika Srbija
4. **V. Nikolić**, M. Ivković, S. Nedeljković, P. Djikanović, INFORMATION RETRIEVAL FOR UNSTRUCTURED TEXT DOCUMENTS: LUCENE SEARCHING, AIIT 2015, Zrenjanin, Republika Srbija
5. **V. Nikolić**, S. Nedeljković, P. Djikanović, INFORMATION RETRIEVAL FOR UNSTRUCTURED TEXT DOCUMENTS: LUCENE INDEXING, EUROBREND 2015
6. **V. Nikolić**, B. Markoski, M. Ivković, K. Kuk, P. Djikanović, Information retrieval for unstructured text documents in Serbian into the crime domain, str. 6., CINTI 2015
7. **V. Nikolić**, R. Dragović, Interoperabilnosti i bezbednost eGovernment Republike Srbije, TELFOR 2014, Kopaonik, Republika Srbija
8. R. Dragović, J. Ivković, D. Dragović, Đ. Klipa, D. Radišić, **V Nikolic**, Sistem za podršku odlučivanju kao podrška za strateško upravljanje državnom upravom, YU INFO 2015, Kopaonik, Republika Srbija
9. D. Randjelović, B. Popović, **V. Nikolić**, S. Nedeljković, Inteligentna pretraga pojmova na primeru policijskih servisa u elektronskoj upravi, Nove informacione tehnologije za analitičko odlučivanje u biološkim, ekonomskim i sociološkim sistemima, (M44), Državni univerzitet u Novom Pazaru, 2014, Republika Srbija
10. P. Djikanović, **V. Nikolić**, D. Sivčević, Nacionalni okvir interoperabilnosti Republike Srbije i servisno orijentisana arhitektura (SOA), YU INFO 2014, Kopaonik, Republika Srbija
11. **V. Nikolić**, J. Protić, P. Djikanović, eGovernment interoperability in the context of European Interoperability Framework (EIF), ICIST 2014, Kopaonik, Republika Srbija
12. **V. Nikolić**, J. Protić, P. Djikanović, G2G INTEGRACIJA MUP-a REPUBLIKE SRBIJE SA PORTALOM E-UPRAVA, ETRAN 2013, Republika Srbija
13. **V. Nikolić**, P. Djikanović, N. Jokić, SISTEM VIDEO NADZORA ZA AUTOMATSKO PREPOZNAVANJE REGISTARSKIH TABLICA MOTORNIM VOZILA, INFOFEST 2013
14. **V. Nikolić**, P. Djikanović, D. Batočanin, eGovernment Republike Srbije - Produženje registracije motornih i priključnih vozila, YU INFO 2013, Kopaonik, Republika Srbija
15. **V. Nikolić**, P. Djikanović, SERVISI MUP-A REPUBLIKE SRBIJE KAO DOPRINOS RAZVOJU E-UPRAVE REPUBLIKE SRBIJE, Kultura polisa, Novi Sad, 2012, Kopaonik, Republika Srbija
16. **V. Nikolić**, S. Radovanović, INTEGRACIJA WEB SERVISA MUP-a REPUBLIKE SRBIJE SA PORTALOM eUPRAVE, INFOTECH 2012, Jahorina, BiH

## **VII ZAKLJUČCI ODNOSNO REZULTATI ISTRAŽIVANJA**

Кандидат мр Војкан Николић је у својој дисертацији решавао проблем одабира најпогодније технологије за израду модела система за брзо проналажење корисних информација из неструктурираних текстуалних докумената на српском језику и његово вредновање. Задатак истраживања био је креирати *question answering* систем базиран на ВоС моделу и алгоритам дела система за класификацију одговора за чланове кривичних дела Кривичног законика. У дисертацији је дат детаљан приказ стања у области предмета истраживања докторске дисертације, наведени су сви важнији пројекти који се баве овом облашћу уз навођење њихових најважнијих карактеристика. Списак коришћене литературе је врло исцрпан и обухвата најважније наслове из ове области. Тачност предложеног система проверена је уз помоћ LUKE графичког едитора компаније Apache који је намењен за анализу резултата брзог претраживања применом библиотеке отвореног кода Lucene. Валидација добијених резултата вршена је поређењем резултата од стране експерата у области права – адвоката. Показало се да су се добијени резултати у груписању одговора на постављени упит у већој мери поклапали са одговорима од стране људских одлука, тј. одговора адвоката. Истраживањем су били обухваћени само они чланови законика који су се односили на телесне повреде у оквиру Кривичног законика. Као репозиторијум за ближе одређивање ВоС модела коришћени су извори (чланци дневних новина Блиц, као и секција питања и одговори портала за пружање бесплатне правне помоћи PRO BONO). На тај начин обезбеђен је адекватан скуп релевантних термина – речи који описују разматране врсте кривичних дела.

Оваквим приступом је превазиђен недостатак примене најкоришћеније статистичке методе TD-IDF у анализи текстуалних докумената. Такође, због специфичности карактеристика српског језика у развоју модела употребљена је техника Н-грам анализа. У процесу мерења сличности при тражењу адекватног одговора вршена је упоредна анализа три мере сличности, при чему је ограничење представљала сама природа кратких текстова, како у постављеним упитима тако и у одговорима кроз чланова закона.

Истраживање које је предмет ове докторске дисертације имало је за циљ да потврди хипотезу да *је могуће креирати ефикасан модел за претраживање неструктурираних података и докумената у е-Управи Републике Србије*. Поред главне хипотезе, постављене су и две помоћне хипотезе а које се односе на груписање неструктурираних докумената на основу тражених података у документима (метадата), као и реализација QA система базираног на ВоС моделу који даје брзе одговоре на упите грађана у вези чланова Кривичног законика Републике Србије. Укупан број кратких текстуалних докумената, који је коришћен у истраживању, био је 540. Главна истраживања за потребе дисертације рађена су на Криминалистичко-полицијској академији, а у току је и реализација примене развијеног модела у центру за Високотехнолошки криминал (ВТК) академије. На основу резултата истраживања са прецизношћу од 75% и тачношћу од 46%, може се рећи да је систем на случају од 540 упита ефикасно елиминисао 130 који нису имали везе са кривичним делима. Разматрањем резултата истраживања која се односе на ефикасност алгорита за класификацију упита у процесу аутоматизације генерисања одговора потребног у случају кратких текстова може се констатовати да постоји значајна уштеда времена у проналажењу адекватних одговора. На основу ових резултата, може се рећи да су потврђене обе помоћне хипотезе. Гледано у целини, на основу добијених одговора система види се да постоји велико подударње са одговорима које су дали експерти – адвокати. На основу претходно изнетих чињеница, може се рећи да је овом дисертацијом постигнут основни циљ истраживања да се може реализовати ефикасан QA систем у склопу е-Управе Републике Србије подржан представљеним концептом претраживања и моделовања неструктурираних података и докумената.

#### **VIII ОЦЕНА НАЧИНА ПРИКАЗА И ТУМАЧЕЊА РЕЗУЛТАТА ИСТРАЖИВАЊА**

Експлицитно навести позитивну или негативну оцену начина приказа и тумачења резултата истраживања.

У дисертацији је јасно представљен предмет и циљ истраживања. Приказана су постојећа решења и јасно исказани резултати дисертације који имају теоријске и практичне доприносе у еволуцији развоја сервиса е-Управе уопште, као интелигентних будућих сервиса е-Управе Републике Србије. Образложене су потребе, осмишљен је и у потпуности је реализован *question answering* систем базиран на ВоС моделу и алгоритама дела система за класификацију упита. У евалуацији су коришћени тренутно расположиви ресурси у виду *on-line* доступних репозиторијума. Рађено је истраживање колико развијени систем доприноси ефикасном напредном претраживању неструктурираних докумената у облику кратких текстова и потврђене су истраживачке хипотезе, што је и публиковано у часопису са импакт фактором као и на међународним конференцијама у земљи и региону. Сви резултати су систематично презентовани у тексту дисертације. Резултати истраживања су прегледно презентовани, детаљно анализирани и дато је њихово тумачење које потврђује исправност постављених циљева и методолошког приступа реализованом истраживању.

## IX КОНАЧНА ОЦЕНА ДОКТОРСKE ДИСЕРТАЦИЈЕ:

Експлицитно навести да ли дисертација јесте или није написана у складу са наведеним образложењем, као и да ли она садржи или не садржи све битне елементе. Дати јасне, прецизне и концизне одговоре на 3. и 4. питање:

1. Да ли је дисертација написана у складу са образложењем наведеним у пријави теме  
Дисертација је у целини написана у складу са образложењем наведеним у пријави теме.

2. Да ли дисертација садржи све битне елементе  
Дисертација садржи све битне елементе који се захтевају по Статуту Факултета и Универзитета у Новом Саду, као и Закона о високом образовању.

3. По чему је дисертација оригиналан допринос науци  
Дисертација представља заокружени процес истраживања који је обухватио дизајн новог QA система базираног на VoC моделу који даје брзе одговоре на упите грађана у вези чланова Кривичног законика Републике Србије. Након уочених трендова у сервисима е-Управа разних Влада, креиран је и развијен систем базиран на савременим интелигентним техникама. Проверене су постављене истраживачке хипотезе и сагледани су правци даљих истраживања и примене развијеног система.

Оригинални доприноси овог рада су:

- Општи преглед могућности и ограничења примене постојећих аналитичких метода за откривање знања из текстуалних докумената.
- Примена општих принципа моделовања и претраживања над реалним подацима у одређеним, постојећим евиденцијама е-Управе Републике Србије.
- Примена постојећих модела у изради *question answering* приступа за ефикаснију примену text mining алгоритама (груписање и кластеровање) над неструктурираним документима.
- Побољшање процене пронађених корисних информација из текстуалних докумената писаних на српском језику.
- Испитивање техника мерења сличности које могу да обезбеде адекватан одговор из екстрахованих делова докумената уместо из целих текстова докумената.
- Успостављање семантички заснованих односа између „корисних“ информација из одређених постојећих репозиторијума е-Управе Републике Србије.
- Оптимизација процеса проналажења конкретног одговора на постављено питање корисника кроз постојећи (будући) сервис е-Управе Републике Србије за пружање услуга брзог одговора грађанима.
- Израда QA система базираног на VoC моделу за претраживање текстуалних докумената на српском језику са могућношћу његове имплементације у области дигиталне форензике.

Резултати су публиковани у 17 референци.

4. Недостаци дисертације и њихов утицај на резултат истраживања  
Утврђено је да дисертација нема недостатака који би утицали на резултате истраживања

<b>X ПРЕДЛОГ:</b>
На основу укупне оцене дисертације, комисија предлаже:
- да се докторска дисертација „МОДЕЛОВАЊЕ И ПРЕТРАЖИВАЊЕ НАД НЕСТРУКТУИРАНИМ ПОДАЦИМА И ДОКУМЕНТИМА У Е-УПРАВИ РЕПУБЛИКЕ СРБИЈЕ“ кандидата мр Војкана Николића <b>прихвати, а кандидату одобри одбрана</b>

У Зрењанину, 13.06.2016.

НАВЕСТИ ИМЕ И ЗВАЊЕ ЧЛАНОВА КОМИСИЈЕ  
ПОТПИСИ ЧЛАНОВА КОМИСИЈЕ

Редовни професор др Миодраг Ивковић ( председник)

Редовни професор др Драгана Глушац (члан)

Доцент др Срђан Попов (члан)

Доцент др Здравко Иванковић (члан)

Ванредни професор др Бранко Маркоски (ментор)

НАПОМЕНА: Члан комисије који не жели да потпише извештај јер се не слаже са мишљењем већине чланова комисије, дужан је да унесе у извештај образложење односно разлоге због којих не жели да потпише извештај.