

ИЗВЕШТАЈ О ОЦЕНИ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ

I ПОДАЦИ О КОМИСИЈИ		
1. Датум и орган који је именовao комисију:		
2. Састав комисије у складу са <i>Правилима докторских студија Универзитета у Новом Саду</i> :		
1. Ивановић др Драган	Редовни професор	Примењене рачунарске науке и информатика, 21.10.2020.
презиме и име	звање	ужа научна област и датум избора
Факултет техничких наука, Универзитет у Новом Саду	Председник	
установа у којој је запослен-а	функција у комисији	
2. Николић др Младен	Ванредни професор	Рачунарство и информатика, 22.02.2022.
презиме и име	звање	ужа научна област и датум избора
Математички факултет, Универзитет у Београду	Члан	
установа у којој је запослен-а	функција у комисији	
3. Рапаић др Милан	Редовни професор	Аутоматика и управљање системима, 07.10.2021.
презиме и име	звање	ужа научна област и датум избора
Факултет техничких наука, Универзитет у Новом Саду	Члан	
установа у којој је запослен-а	функција у комисији	
4. Сливка др Јелена	Ванредни професор	Примењене рачунарске науке и информатика, 10.07.2020.
презиме и име	звање	ужа научна област и датум избора
Факултет техничких наука, Универзитет у Новом Саду	Члан	
установа у којој је запослен-а	функција у комисији	
5. Ковачевић др Александар	Редовни професор	Примењене рачунарске науке и информатика, 27.01.2022.
презиме и име	звање	ужа научна област и датум избора
Факултет техничких наука, Универзитет у Новом Саду	Ментор	
установа у којој је запослен-а	функција у комисији	
II ПОДАЦИ О КАНДИДАТУ		
1. Име, име једног родитеља, презиме: Александар, Драгана, Каплар		

2. Датум рођења, општина, држава:
05.05.1991., Бијељина, Босна и Херцеговина
3. Назив факултета, назив претходно завршеног нивоа студија и стечени стручни/академски назив:
Факултет техничких наука Универзитета у Новом Саду, Рачунарство и аутоматика, Дипломирани инжењер електротехнике и рачунарства - мастер.
4. Година уписа на докторске студије и назив студијског програма докторских студија:
2015, Рачунарство и аутоматика

III НАСЛОВ ДОКТОРСKE ДИСЕРТАЦИЈЕ:

Аутоматско издвајање семантике из медицинских докумената на српском језику

IV ПРЕГЛЕД ДОКТОРСKE ДИСЕРТАЦИЈЕ:

Навести кратак садржај са назнаком броја страница, поглавља, слика, схема, графикона и сл.

Докторска дисертација припада научној области електротехника и рачунарство, односно ужој научној области примењене рачунарске науке и информатика. Написана је на српском језику на 155 страна Б5 формата латиничним писмом. Главни део дисертације садржи 8 поглавља, уз додатне сегменте за апстракт на енглеском језику, резиме рада на српском језику, литературу, биографију, прилог, индекс слика, индекс табела и план третмана података. Дисертација садржи 29 слика, 37 табела и 128 навода из литературе. Кључна документацијска информација написана је на српском и енглеском језику.

Докторска дисертација се састоји од следећих поглавља:

1. Увод
2. Теоријске основе
3. Преглед актуелног стања у области
4. Корпус златног стандарда
5. Модел система за препознавање именованих ентитета
6. Експериментални резултати
7. Дискусија
8. Закључак

V ВРЕДНОВАЊЕ ПОЈЕДИНИХ ДЕЛОВА ДОКТОРСKE ДИСЕРТАЦИЈЕ:

Предмет истраживања ове докторске дисертације је развој система за препознавање именованих ентитета из медицинских документа написаних на српском језику. Препознавање именованих ентитета припада научној области обраде природног језика. У оквиру истраживања коришћени су медицински извештаји са Клинике за нефрологију Универзитетског клиничког центра Србије.

Докторска дисертација је организована у осам поглавља.

У првом, уводном, поглављу дат је опис мотивације истраживања са јасно дефинисаним истраживачким проблемима, хипотезом и циљевима истраживања.

У другом поглављу су описане потребне теоријске основе везане за области које обухвата ова докторска дисертација. Описане су теоријске основе приступа за препознавање именованих ентитета. Детаљно су приказани приступи чија се методологија базира на машинском учењу. Посебна пажња посвећена је методологијама које користе моделе дубоког учења, јер они представљају најсавременије приступе у области обраде природног језика.

У трећем поглављу дат је преглед тренутног стања у области са фокусом на препознавање именованих ентитета у медицинском домену. Дат је преглед историје настанка и примене алгоритама за препознавање именованих ентитета у општем као и у медицинском домену. У самом прегледу акценат је стављен на типове коришћених алгоритама и на остварене резултате. Посебна пажња је посвећена истраживањима за препознавање именованих ентитета у медицинским документима на српском језику.

У четвртном поглављу описан је процес формирања корпуса златног стандарда који представља један од резултата ове докторске дисертације. Детаљно су описани процес формирања анотационе шеме, процес анотације и евалуације слагања анотатора. Приликом формирања анотационе шеме дат је кратак опис анотационих шема из сродних истраживања, и дат је опис категорија (догађаји, временски изрази, вредности) и значаја класа именованих ентитета у развијеној анотационој шеми. Након тога су дати резултати анализе поузданости анотатора и приказане су све потребне статистичке карактеристике аотираног корпуса.

У петом поглављу детаљно је описана архитектура развијеног система за препознавање именованих ентитета у медицинским документима. Архитектура система је дата као комбинација више различитих типова модела машинског учења организованих у модуле. Након представљања опште архитектуре система, редом су описани модули система: модул за претпроцесирање, модул за екстракција особина, модул за формирање вектора значења речи, и модул за детекцију именованих ентитета.

У шестом поглављу приказани су резултати система. Важан део овог поглавља је и опис методологије за евалуацију система која укључује експерименталну поставку и мере перформанси. Детаљно су анализирани резултати коришћених модела машинског учења, односно модула система. У наставку поглавља су приказани резултати развијеног система за препознавање именованих ентитета употребом стандардних мера за евалуацију класификационих система на нивоу индивидуалних токена као и на нивоу именованих ентитета.

У седмом поглављу дата је дискусија. Дискутовани су резултати развијеног система за препознавање именованих ентитета, предност и мане модела дубоког учења, као и утицај различитих приступа приликом обучавања модела дубоког учења на перформансе обученог модела. У оквиру дискусије налази се детаљна анализа грешака система и идентификованих ограничења истраживања.

У осмом поглављу наведен је закључак ове дисертације. Поред осврта на кључне детаље истраживања, наведене су смернице за даља истраживања и могућа побољшања система.

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 складу са *Правилима докторских студија Универзитета у Новом Саду* који је повезан са садржајем докторске дисертације. У случају радова прихваћених за објављивање, таксативно навести називе радова, где и када ће бити објављени и приложити потврду уредника часописа о томе.

M21 – Врхунски међународни часопис

1. **Kaplar, A.,** Stošović, M., Kaplar, A., Brković, V., Naumović, R., & Kovačević, A., Evaluation of clinical named entity recognition methods for Serbian electronic health records, *International Journal of Medical Informatics*, 164 (2022), pp. 104805, ISSN: 1386-5056 , DOI: 10.1016/j.ijmedinf.2022.104805

M33 – Саопштење са међународног скупа штампаног у целини

1. **Kaplar, A.,** Aleksić, A., Stošović, M., Naumović, R., Brković, V., Kovačević, A., Evaluating String Distance Metrics for Approximate Dictionary Matching: A Case Study in Serbian Electronic Health Records, *In: Konjović, Z., Zdravković, M., Trajanović, M. (Eds.) ICIST 2019 Proceedings*, 2019, pp.135-137, ISBN: 978-86-85525-24-7
2. **Kaplar, A.,** Simić, M., Kovačević, A., Artificial General Intelligence Approach for Reasoning in Clinical Decision Support, *In: Zdravković, M., Konjović, Z., Trajanović, M. (Eds.) ICIST 2017 Proceedings*, 2017, pp.271-274, ISBN: 978-86-85525-19-3

VII ЗАКЉУЧЦИ ОДНОСНО РЕЗУЛТАТИ ИСТРАЖИВАЊА:

У докторској дисертацији је представљено истраживање из области обраде природног језика. Препозната је потреба предложеног система за медицинске институције у Србији. Омогућена је аутоматско екстраховање и директно коришћење информација из неструктурираних делова медицинских докумената, као и коришћење екстрахованих именованих ентитета за даља истраживања у области медицинске обраде природног језика. Представљена методологија није доменски специфична и може се применити у свим медицинским установама у Србији. Описана методологија верификована је имплементацијом прототипа система и евалуацијом истог.

Резултати истраживања су:

- анализа постојећих система за препознавање именованих ентитета у медицинском домену
- формирање методологије за препознавање именованих ентитета у медицинским документима
- формирање анотираног корпуса медицинских докумената на српском језику
- развијање алата и ресурса за детекцију реченица у медицинским документима на српском језику
- креирање модела за препознавање именованих ентитета из медицинских докумената на српском језику
- имплементација прототипа система за препознавање именованих ентитета из медицинских докумената на српском језику
- дискусија резултата и анализа грешака као основа за даљи развој система за препознавање именованих ентитета у медицинским документима на српском језику

VIII ОЦЕНА НАЧИНА ПРИКАЗА И ТУМАЧЕЊА РЕЗУЛТАТА ИСТРАЖИВАЊА:

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

У докторској дисертацији је на јасан и прегледан начин представљено истраживање у области обраде природног језика. Изложене су теоријске основе релевантне за истраживање, извршена је анализа тренутног стања у овој области, представљена методологија, извршени експерименти и приказани резултати истраживања. Извршена је дискусија резултата и изведени су закључци који сумирају доприносе, наводе ограничења истраживања и могуће правце даљих истраживања. Резултати истраживања су приказани прегледно и систематично, помоћу табела и слика које олакшавају њихово тумачење. Извршена је софтверска провера докторске дисертације на плагијаризам у библиотеци ФТН софтвером за детекцију плагијаризма *iThenticate*.

У складу са наведеним Комисија ПОЗИТИВНО оцењује начин приказа и тумачења резултата истраживања.

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. Навести нумеричке податке о резултатима провере оригиналности рада и дати текстуално образложење.

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

Да. Докторска дисертација је написана у складу са образложењем наведеним у пријави теме.

2. Да ли дисертација садржи све битне елементе?

Да. Дисертација садржи све битне елементе.

3. По чему је дисертација оригиналан допринос науци?

Оригинални научни допринос докторске дисертације чини доказ полазне хипотезе на основу експерименталног и теоријског истраживања. На основу предложеног модела, извршених експеримента и добијених резултата доказна је хипотеза да је могуће применити савремене технике, засноване на машинском учењу, за аутоматско препознавање именованих ентитета из медицинских докумената на српском језику. Методологија и сазнања која су представљена у овој докторској дисертацији пружају преко потребне основе за даљи рад на аутоматском препознавању именованих ентитета у медицинском домену за српски језик који је у овој области недовољно истражен и има недостатак језичких ресурса.

Резултати дисертације су објављени у врхунском међународном часопису (M21).

4. Који су недостаци дисертације и какав је њихов утицај на резултат истраживања?

Дисертација нема битне недостатке који утичу на резултате истраживања.

5. Образложење резултата провере оригиналности рада (нумерички и наративно):

Провером рада на плагијаризам помоћу програмског пакета iThenticate утврђен је проценат преклапања од 3%, при чему се подударане односи на терминолошке и методолошке фразе. Такође, проценат преклапања са сваким наведеним извором је мањи од 1%, те Комисија констатује да је кандидат адекватно цитирао коришћену литературу и да оригиналност добијених научних резултата и дискусије није упитна.

X ПРЕДЛОГ:
На основу наведеног, комисија предлаже:
а) да се докторска дисертација прихвати, а кандидату одобри одбрана;

Место и датум: Нови Сад, 15.07.2024.

1. др Драган Ивановић, Редовни професор
_____, председник
2. др Младен Николић, Ванредни професор
_____, члан
3. др Милан Рапаић, Редовни професор
_____, члан
4. др Јелена Сливка, Ванредни професор
_____, члан
5. др Александар Ковачевић, Редовни професор
_____, 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 и да исти потпише.