

НАСТАВНО-НАУЧНОМ ВЕЋУ

Предмет: Подобност теме и кандидата **Браниславе Цвијетић**, магистра техничких наука област аутоматика и информатика, за израду докторске дисертације **„Откривање структура података употребом интегрисаног алгоритма заснованог на техникама машинског учења”** (на енглеском *„Discovering data structures using an integrated algorithm based on machine learning techniques”*)

Одлуком бр. 920-31 од 14.05.2024. године, именовани смо за чланове Комисије за оцену подобности теме и кандидата Браниславе Цвијетић за израду докторске дисертације и научне заснованости теме **„Откривање структура података употребом интегрисаног алгоритма заснованог на техникама машинског учења”** (на енглеском *„Discovering data structures using an integrated algorithm based on machine learning techniques”*)

На основу материјала приложеног уз Захтев кандидата, Комисија подноси следећи

ИЗВЕШТАЈ

1. Подаци о кандидату

1.1. Биографски подаци

Бранислава Цвијетић је рођена 26.01.1978. године у Сарајеву, Босна и Херцеговина. Електротехнички факултет Универзитета у Источном Сарајеву, уписала је 1996. године, а 2002. године је дипломирала на одсеку аутоматика и електроника на тему „Пројектовање и имплементација ПИД контролера” код проф. др Милице Б. Наумовић.

Постдипломске студије, одсек аутоматика и информатика, уписала је 2003. године на Електротехничком факултету Универзитета у Источном Сарајеву. Магистрирала је 2013. године код проф. др Слободанке Ђорђевић – Кајан на тему „Пројектовање и имплементација сигурности код система база података”.

Докторске студије уписала је 2016. године на модулу рачунарска техника и информатика Електротехничког факултета Београд и положила све испите са просечном оценом 10.00. Упоредно са редовим радним задацима, дефинисаним радним местом на којем је ангажована, бави се истраживањем проблема у области вештачке интелигенције, односно примене машинског учења у решавању проблема у домену рада.

Од 2003. године до 2006. године радила је на радном месту виши стручни сарадник – програмер у Републичком заводу за статистику Републике Српске. Од 2006. до 2007. радила је на радном месту стручни сарадник за базе података у Централној изборној комисији БиХ. Од 2007. до 2010. године радила је у Државној агенцији за истраге и заштиту. Од 2010.

године запослена је у Агенцији за статистику БиХ на радном месту стручни саветник за пројектовање и програмирање.

У периоду од 15.01. - 15.06.2018. била је на стручном усавршавању у Заводу за статистику Републике Грчке у оквиру Програма обуке за статистичаре у Еуростату и државама чланицама Европске уније – Вишекорисничког програма статистичке сарадње. Стручно усавршавање је било на тему „Рад на експерименталним пројектима везаним за Велике податке”.

1.2. Стечено научноистраживачко искуство

Кандидат је докторске академске студије на изборном подручју Електротехника и рачунарство на ЕТФ-у уписао 2016/2017. школске године, а потом 2023/2024. године. Од тренутка уписа до тренутка подношења Захтева за одобрење докторске дисертације положио је све испите (са просечном оценом 10,0) и обавио све обавезе предвиђене докторским академским студијама на изборном подручју Електротехника и рачунарство. Списак положених испита и обављених обавеза је дат у следећој табели:

бр.	Назив испита / обавезе	Оцена	ЕСПБ	Датум
1	Вештачка интелигенција и експертски системи (признат испит)	10	9	02.07.2018.
2	Софтверски дефинисане мреже (признат испит)	10	9	15.01.2018.
3	Студијски истраживачки рад I	-	24	02.04.2024.
4	Студијски истраживачки рад II	-	24	02.04.2024.

Бранислава Цвијетић се бави развојем метода за откривање структура података коришћењем техника машинског учења. Циљ истраживања је да се предложи алгоритам и развије аутоматизовано софтверско решење за откривање структура унутар података код којих је дошло до нарушавања структуре, а које је модулarno са становишта броја функција за рачунање сличности између структура необележеног скупа података. Такође, решење треба да омогући проширивање са новим алгоритмима машинског учења, могућношћу промене одређених параметара у погледу дефинисања оквира података за истраживање, додавања правила за одређивање врсте података и сл. Садржај генерисаних ентитета даље се може машински обрадити у релационим базама података и као такав послужити у процесу стицања знања и доношења пословних одлука, што представља основни значај предложеног истраживања. Додатни значај огледа се у чињеници да ће алгоритам предвидети могућност сажимања великог скупа података и омогућити примену алгоритама машинског учења уз уштеду времена и рачунарских ресурса, а решење неће бити ограничено доменом примене. За потребе рада биће креирани скупови података са нарушеном структуром на основу јавно доступних података. Бранислава Цвијетић је предложила решење за откривање структура података употребом интегрисаног алгоритма заснованог на техникама машинског учења.

Бранислава Цвијетић до сада је објавила два рада у међународном часопису (један часопис је са СЦИ листе, док је други ван СЦИ листе) као први аутор и пет радова на међународној конференцији (пет као први аутор).

Дана 31.05.2024. године одржана је јавна усмена одбрана теме докторске дисертације (докторски испит) пред Комисијом у саставу: др Марија Пунт, ванредни професор са Електротехничког факултета; др Синиша Влајић, редовни професор са Факултета организационих наука; др Зоран Чича, редовни професор са Електротехничког факултета. Оцена Комисије је да је кандидат Бранислава Цвијетић задовољила све критеријуме и одбранио предложену тему докторске дисертације.

У наставку је списак радова који показују резултате досадашњег рада и афинитета који су у складу са предложеном дисертацијом кандидата Браниславе Цвијетић.

Радови у међународним часописима (са СЦИ листе):

B. Cvijetić, Z. Radivojević, Restoration of Data Structures Using Machine Learning Techniques, IEEE ACCESS, Vol. 11, pp. 113077 - 113099, Oct, 2023, DOI: 10.1109/ACCESS.2023.3323846, Impact factor: 3.9 (M22)

Радови у међународним часописима (ван СЦИ листе):

B. Cvijetic, Z. Radivojevic, Application of Machine Learning in the Process of Classification of Advertised Jobs, International Journal of Electrical Engineering and Computing, Vol. 4, No. 2, pp. 93 - 100, Dec, 2020

Радови на домаћим конференцијама:

B. Cvijetić, Z. Radivojević, Primjena mašinskog učenja u procesu klasifikacije oglašених radnih mjesta, XIV међународни симпозијум INFOTЕH-JAHORINA 2020, pp. 146 - 151, Jahorina, Istočno Sarajevo, Republika Srpska, Bosna i Hercegovina, Mar, 2020

B. Cvijetić, Z. Radivojević, Primjena Weka softvera i konvolucijskih neuronskih mreža u procesu klasifikovanja digitalnih fotografija, XVIII међународни симпозијум INFOTЕH-JAHORINA 2019, pp. 224 - 229, Jahorina, Istočno Sarajevo, Republika Srpska, Bosna i Hercegovina, Mar, 2019

B. Cvijetić, Otvaranje javnih podataka u Bosni i Hercegovini, XIV међународни симпозијум INFOTЕH-JAHORINA 2015, pp. 670 - 674, Jahorina, Istočno Sarajevo, Republika Srpska, Bosna i Hercegovina, Mar, 2015

B. Cvijetić, D. Avram, M. Bratić, Upotreba kriptografije za sigurnost baza podataka, VII међународни симпозијум INFOTЕH-JAHORINA 2008, Jahorina, Istočno Sarajevo, Republika Srpska, Bosna i Hercegovina, Mar, 2008

M. Naumović, D. Avram, B. Cvijetić, Neuronske mreže u sistemima upravljanja kroz primere, IV међународни симпозијум INFOTЕH-JAHORINA 2005, Jahorina, Istočno Sarajevo, Republika Srpska, Bosna i Hercegovina, Mar, 2005

1.3. Оцена подобности кандидата за рад на предложеној теми

Приликом писања овог Извештаја, Комисија је размотрила све претходно наведене чињенице и резултате у досадашњем раду Браниславе Цвијетић. Са посебном пажњом су размотрена искуства у истраживачком раду и узимајући све наведено, Комисија је стекла увид у њену способност за научно истраживачки рад и рад на предложеној теми докторске дисертације, која је, такође, препозната и вреднована и од стране научне заједнице.

2. Предмет, циљ и значај истраживања

Са напретком информационо-комуникационих технологија, количина доступних и похрањених електронски података значајно се повећала током последњих година. *International Data Corporation (IDC)*, као једна од водећих компанија за прикупљање глобалних информација о: тржишту, подацима, као и догађајима везанима за информационе технологије и телекомуникације, на основу прогнозе од 2021. до 2026. године, предвиђа да ће подаци расти по сложеној стопи раста (енгл. *compound annual growth rate CAGR*) од 21,2% како би достигли више од 221.000 ексабајта до 2026. године. Ти подаци ће укључивати и структуриране и неструктуриране податке, али неструктурирани подаци у великој вечини доминирају, чинећи више од 90% података креираних сваке године [1].

Сви подаци који се могу сачувати, прегледати и обрадити у облику фиксног формата називају се структурираним подацима (енгл. *Structured data*). Наиме, структурирани подаци су организовани у семантичке ентитете где су слични ентитети груписани у виду релација, а ентитети исте групе имају исти опис, тј. исте атрибуте. Описи за све ентитете у некој групи чине схему. Атрибути обично имају предефинисан формат и дужину, а обично сви следе неки предефинисани редослед. Пример структурираних података укључује различите релационе податке који су смештени у релационим базама података. Скупови података који не морају нужно имати неки дефинисан формат или следити неко правило по питању садржаја података потпадају у групу неструктурираних података (енгл. *Unstructured data*). У неструктуриране податке спадају текстуални подаци, видео и аудио записи, као и фотографија. Подаци који садрже оба наведена облика података називају се полуструктурирани (енгл. *Semi-structured data*) подаци. Насупрот структурираним подацима, код полуструктурираних података, ентитети исте групе могу да имају различите атрибуте, а редослед атрибута није нарочито важан, док се за означавање самих атрибута користе одговарајући маркери или лабеле. Примери полуструктурираних података су емаил, језик за означавање хипертекста (енгл. *Hypertext Markup Language HTML*), формати за размену података (нпр. *Electronic data interchange EDI*, и *JavaScript Object Notation JSON*).

Поред чињенице да постоји огромна количина неструктурираних података, као и полуструктурираних података, то доводи до вишеструких изазова за управљањем таквим подацима у погледу добијања додатне вредности из података. Извлачење информација (енгл. *Information extraction - IE*) из полуструктурираних и неструктурираних података је област која се стално истражује пошто постоји потреба да се из таквих података открију шеме података (енгл. *schema discovery*), као и подаци повезани са тим шемама.

У радовима [2]–[5] анализиран је утицај приступа откривању шема података, техника за откривање шема података, карактеристика процедура откривања, формата улазних и излазних података, као и квалитета резултата. Приступи откривању шеме података се састоје од неколико корака, у којима се информације у вези са шемом издвајају из оригиналног улазног скупа информација шеме. Ове информације се могу издвојити из РДФ-а, ОЕМ-а или других типова података. Приступи откривању шеме могу се груписати у три категорије на основу основних алгоритаамских концепата који стоје иза методе откривања шеме. Ове категорије укључују имплицитно откривање шема, експлицитно обogaћивање постојећих шема и откривање структурних образаца унутар постојећих шема.

Приступи имплицитног откривања шеме покушавају да открију шему из извора података без потребе за додатним информацијама о шеми. Ова процедура има за циљ да издвоји нове информације из скупа података анализом ентитета и веза између елемената. Ови приступи почињу анализу са празном шемом и покушавају да окарактеришу овај скуп података на два начина: груписање сличних инстанци и сличних путања [6]–[10].

Приступи који користе експлицитно обogaћивање постојећих шема почињу са постојећим шемама и исказима о шеми који могу да допуне или обогате постојеће шеме. Њихов циљ је да генеришу нове типове користећи постојеће типове или да генеришу друге шеме. Ови приступи се могу класификовати у две категорије. Први приступ је заснован на техникама

машинског учења, док су остали приступи засновани на статистичкој анализи података. Ове технике првенствено користе повезаност између инстанци [11]–[13].

Процедуре за откривање структурних образаца унутар постојећих шема почињу идентификацијом свих могућих структурних образаца (верзија) ентитета у датом скупу података познате шеме. Ова процедура има за циљ да анализира однос заједничког појављивања између својстава у скупу података пре него да открије типове и додатне шеме, као у експлицитном обogaћивању постојећих шема. Различити скуп својстава описује сваку инстанцу унутар скупа података. Скуп својстава који описује инстанцу представља њен структурни образац. Ови приступи се могу класификовати у две категорије на основу обима препознавања образаца. Први је заснован на тачном препознавању образаца. Насупрот томе, други приступ је заснован на приближном препознавању образаца. Радови који описују тачне процедуре за препознавање образаца дати су у [14]–[20], док су они који описују приближне приступе откривању образаца дати у [21] и [22].

Посебна врста неструктурираних података су подаци за које се се поуздано зна да поседују структуру, али која је из било ког разлога нарушена. Овакав скуп података у свом саставу може садржавати различите верзије исправних структура једног или више ентитета, а које су мењане током времена (нпр. замена у редоследу атрибута, броју атрибута и сл.). Такође може садржавати и нарушене структуре ентитета које су настале при експорту структурираних података услед коришћења различитих делимитера за раздвајање атрибута ентитета, као и новог реда. Осим наведених особина којима се описује овакав скуп података, проблем настаје и због саме хетерогености података. Подаци са великом флукуацијом и различитим типом података називају се хетерогени подаци. Хетерогени скупови структурираних података могу садржавати информације о датумима или временским интервалима, експлицитне нумеричке податке (нпр. година), кодиране нумеричке податка (нпр. 0 или 1 за присуство или одсуство), категоријске податке (нпр. земље) и текстуална поља (нпр. кратки опис). Такође, код оваквих скупова података постоји проблем који се огледа у чињеници да за неки атрибут могу недостајати вредности, тј. вредност таквих атрибута могу да остану непопуњене или попуњене неком подразумеваном вредношћу. Ови скупови података могу спадати у категорију великих података (енгл. *Big data*), тешко обрадиви са доступним рачунарским ресурсима, а најчешће не садрже мета податке. За скуп података, са претходно наведеним карактеристикама, може се рећи да је неструктуриран, али да се поуздано зна да је некад имао структуру. Предмет истраживања у склопу ове докторске дисертације представља развој проширеног система за експлицитно откривање структура унутар скупа података за који се поуздано зна да поседују структуру, али која је из било ког разлога нарушена.

Циљ истраживања, у склопу ове дисертације, је да се предложи алгоритам и развије аутоматизовано софтверско решење за откривање структура унутар података код којих је дошло до нарушавања структуре, а које је модуларно са становишта броја функција за рачунање сличности између структура необележеног скупа података. Такође, решење треба да омогући проширивање са новим алгоритмима машинског учења, могућношћу промене одређених параметара у погледу дефинисања оквира података за истраживање, додавања правила за одређивање врсте података и сл. Наиме, софтверско решење треба да буде аутоматизовано на начин да на улазу прихвати скуп података за који је познато да је раније био структуриран, те да није познат делимитер који је коришћен за раздвајање атрибута, а као излаз да генерише ентитете. Садржај генерисаних ентитета даље се може машински обрадити у релационим базама података и као такав послужити у процесу стицања знања и доношења пословних одлука, што представља основни значај предложеног истраживања. Додатни значај огледа се у чињеници да ће алгоритам предвидети могућност сажимања великог скупа података и омогућити примену алгоритма машинског учења уз уштеду времена и рачунарских ресурса [23]. Такође, значај овог истраживања је и у томе што решење неће бити ограничено доменом примене, а рад доменских стручњака ће бити минималан. Поред свега наведеног, значајан резултат истраживања је тај да ће, а за потребе

верификације предложеног алгоритама и омогућавања даљих истраживања, бити креирани скупови података са нарушеном структуром коришћењем јавно доступних података.
Списак референци везаних за тезу:

- [1] E. Burgener, and J. Rydning, "High Data Growth and Modern Applications Drive New Storage Requirements in Digitally Transformed Enterprises," IDC Doc. #US49359722, July 2022.
- [2] K. Menouer, N. Kardoulakis, G. Troullinou, Z. Kedad, D. Plexousakis, and H. Kondylakis, "A survey on semantic schema discovery," *The VLDB Journal*, vol. 31, pp. 675-710, 2022.
- [3] F. Bai, J. Kang, G. Stanovsky, D. Freitag, and A. Ritter, "Schema-Driven Information Extraction from Heterogeneous Tables," *ArXiv*. /abs/2305.14336, 2023.
- [4] M. Hassan, K. Shaukat, D. Niu, S. Mahreen, Y. Ma, M. Mubashir, and Z. Xiuyang, "An Overview of Schema Extraction and Matching Techniques," 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Xi'an, China, pp. 1290-1294, 2018.
- [5] E. Koci, M. Thiele, O. Romero, and W. Lehner, "Table identification and reconstruction in spreadsheets," In *International Conference on Advanced Information Systems Engineering*, pp. 527 - 541, Springer, 2017
- [6] E. Rahm, and P. Bernstein, "A survey of approaches to automatic schema matching," *The VLDB Journal*, vol. 10, pp. 334-350, 2001.
- [7] K. Menouer, and Z. Kedad, "Schema discovery in RDF data sources," *Conceptual Modeling-34th International Conference, ER 2015*, vol. 9381, pp. 481-495, 2015.
- [8] M. Kirchberg, E. Leonardi, Y. Tan, S. Link, R. Ko, and B. Lee, "Formal Concept Discovery in Semantic Web Data," *International Conference on Formal Concept Analysis, ICFCA 2012*, vol. 7278, pp. 164-179, 2012.
- [9] P. Li, Y. Gong, and C. Wang, "Schema Extraction on Semi-structured Data," *arXiv:2012.08105*, 2021.
- [10] S. Jain, A. Buitleir, and E. Fallon, "A Review of Unstructured Data Analysis and Parsing Methods," *2020 International Conference on Emerging Smart Computing and Informatics (ESCI)*, pp. 164-169, 2020.
- [11] R. Bouhamoum, Z. Kedad, and S. Lopes, "Schema Discovery in Large Web Data Sources," *International Conference on Big Data and Cyber-Security Intelligence, BDSCIntell*, 2022.
- [12] A. Ezugwu, A. Ikotun, O. Oyelade, L. Abualigah, J. Agushaka, C. Eke, and A. Akinyelu, "A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects," *Engineering Applications of Artificial Intelligence*, vol. 110, 2022.
- [13] A. S. Shirkhorshidi, S. Aghabozorgi, T.Y. Wah, and T. Herawan, "Big Data Clustering: A Review," *International Conference on Computational Science and Its Applications, ICCSA 2014*, vol. 8583, pp. 707-720, 2014.
- [14] H. Dong, S. Liu, S. Han, Z. Fu, and D. Zhang, "TableSense: Spreadsheet Table Detection with Convolutional Neural Networks," *AAAI*, vol. 33, pp. 69-76, 2019.
- [15] C. Christodoulakis, E. Munson, M. Gabel, A. D. Brown, and R. J. Miller, "Pytheas: Pattern-based Table Discovery in CSV Files," *PVLDB*, vol. 3, pp. 2075-2089, 2020
- [16] G. Burg, A. Nazábal, and C. Sutton, "Wrangling messy CSV files by detecting row and type patterns," *Data Mining and Knowledge Discovery*, vol. 33, pp. 1799-1820, 2019.
- [17] L. Jiang, G. Vitagliano, and F. Naumann, "Structure detection in verbose CSV files," In *Proceedings of the International Conference on Extending Database Technology (EDBT)*, pp.193-204, 2021.
- [18] E. Koci, M. Thiele, O. Romero, and W. Lehner, "Table identification and reconstruction in spreadsheets," In *International Conference on Advanced Information Systems Engineering*, Springer, pp. 527-541, 2017.

- [19] J. C. Roldán, P. Jiménez, and R. Corchuelo, "On extracting data from tables that are encoded using HTML," *Knowledge-Based Systems*, vol. 190, 2020.
- [20] G. Yuan, J. Lu, Z. Yan, and S. Wu, "A Survey on Mapping Semi-Structured Data and Graph Data to Relational Data," *ACM Computing Surveys*, vol. 55, pp 1-38, 2023.
- [21] A. O. Shigarov and A. A. Mikhailov, "Rule-based spreadsheet data transformation from arbitrary to relational tables," *Information Systems*, vol. 71, pp. 123-136, 2017.
- [22] H. Elmeleegy, and J. Madhavan, "Harvesting relational tables from lists on the web," *VLDB*, vol. 21, pp. 1078-1089, 2009.
- [23] A. S. Shirkhorshidi, S. Aghabozorgi, T.Y. Wah, and T. Herawan, "Big Data Clustering: A Review," *Computational Science and Its Applications – ICCSA 2014*, Vol. 8583, 2014.

3. Полазне хипотезе

Полазне хипотезе у овом истраживању су:

- Могуће је предложити алгоритам, заснован на коришћењу техника машинског учења, којим се рестаурира структура код података за које са се поуздано зна да поседују структуру, али која је из било ког разлога нарушена.
- Предложени алгоритам моћи ће да даје боље резултате у односу на постојећа решења која су примењива при решавању проблема рестаурирања структуре података са нарушеном структуром.
- Могуће је проширити предложени алгоритам скупом правила и функција за откривање скаларних типова атрибута унутар идентификоване структуре или више структура.
- Могуће је проширити предложени алгоритам скупом правила и функција за откривање векторских типова података.

4. Научне методе истраживања

У овом истраживању биће примењене дескриптивне, експерименталне и комперативне методе.

У првом делу истраживања биће извршен детаљан преглед постојећих решења из отворене литературе, а која се бави проблемом детекције структура из података са нарушеном структуром.

Након тога ће бити анализирани постојећи репозиторијуми података у циљу проналажења адекватних тестних података, односно података који садрже податке који имају један или више референцијалних ограничења, као и векторске типове података фиксне и променљиве дужине.

Потом ће бити анализирана међузависност између података унутар структура, коришћењем постојећих статистичких метода и алгоритама машинског учења. У складу са добијеним резултатима претходних анализа, биће предложен алгоритам за детекцију структура у подацима са нарушеном структуром.

У наредном делу истраживања биће имплементиран прототип алгоритма за откривање структура у подацима са нарушеном структуром, а који се заснива на алгоритмима машинског учења.

У експерименталном делу, над идентификованим улазним подацима, биће извршена анализа резултата примене алгоритама машинског учења, као и детаљно поређење резултата у погледу перформанси и комплексности алгоритама предложених решења. Наиме, на основу резултата добијених током експеримента, применом компаративних метода, биће извршен одабир техника и алгоритама машинског учења.

5. Очекивани научни допринос

Основни допринос раду јесте развој алгоритама за рестаурирање структура код података са нарушеном структуром, односно откривање ентитета и односа међу атрибутима ентитета.

Такође се очекује да ће истраживање дати и следеће научне доприносе:

- Евалуација постојећих алгоритама машинског учења који се могу применити у домену проблема откривања структуре података са нарушеном структуром, као и предлога њихове класификације са аспекта могућности примене у фазама развоја софтверског решења.
- Предложен нови алгоритма који користи технике машинског учења.
- Креирани нови репозиторијуми података са нарушеном структуром на основу постојећих јавно доступних скупова података у отвореној литератури и индустрији. Скупови података ће садржати модификоване податке, тј. податке где је вештачки направљено нарушавање структуре података (на пример, замена атрибута, уклањање или додавање атрибута и сл.).
- Модуларно софтверско решење развијено сходно предложеном алгоритму за откривање структуре унутар података са нарушеном структуром и прошириво са новим модулима за откривање зависности између података, као и њиховог груписања.
- Дизајнирани експерименти погодни за евалуацију перформанси предложеног алгоритма и развијеног софтверског решења.

Сви добијени резултати биће дати на увид јавности кроз публикације на научно-стручним скуповима и часописима.

6. План истраживања и структура рада

Планирано је да се истраживање спроведе у седам (7) фаза.

- Прва фаза обухвата анализу доступне релевантне научно-стручне литературе, као и анализе јавно доступних скупова података који ће се користити у истраживању. У првој фази биће креирани тестни скупови података за потребе процене добијеног решења.
- Друга фаза обухвата анализу постојећих алгоритама и решења за откривање структура унутар података са нарушеном структуром.
- Трећа фаза обухвата предлог алгоритма, функција и метода за препознавање структура података на основу типова података, њиховог распореда, учесталости појављивања одређеног распореда података, као и развоја софтверског решења заснованог на предложеном алгоритму. У трећој фази биће креирано софтверско решење.
- Четврта фаза обухвата тестирање и евалуацију предложеног алгоритма за препознавање структура ентитета. У овој фази се врши и процена да ли софтверско решење исправно груписало ентитете и припадајуће атрибуте на нивоу ентитета тамо где је дошло до нарушавања структуре због непознатог делимитера.
- Пета фаза обухвата тестирање и евалуације предложеног алгоритма за препознавање стварних вредности атрибута унутар препознатих структура и њиховог даљег груписања уколико се ради о погрешно идентификованим стварним вредностима атрибута скаларног или векторског типа података.
- У шестој фази обавља се интегрално тестирање и провера рада предложеног алгоритма и развијеног софтверског решења над креираним скуповима података. Наиме, врши се

провера да ли и колико успешно, развијено софтверско решење може препознати ентитете и њихове припадајуће атрибуте.

- Седма фаза се односи на анализу добијених резултата и доношења одговарајућих закључака.

Уколико резултати и закључци из седме фазе буду указивали да предложени алгоритам не даје очекиване резултате, приступиће се понављању одговарајућих фаза и извршити кориговање корака рада сходно новим претпоставкама.

7. Закључак и предлог

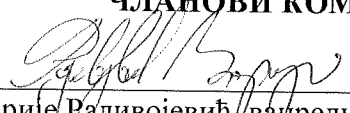
На основу приложене документације и стеченог целокупног утиска о теми докторске дисертације и досадашњег научног рада кандидата **Браниславе Цвијетић**, магистра техничких наука, Комисија је дошла до следећих закључака:

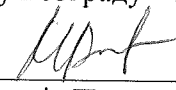
1. Кандидат испуњава све формалне и суштинске потребне услове прописане Законом о високом образовању, Статутом Универзитета у Београду и Статутом Електротехничког факултета Универзитета у Београду, за израду докторске дисертације.
2. Предложена тема **Откривање структура података употребом интегрисаног алгоритма заснованог на техникама машинског учења** (на енглеском „*Discovering data structures using an integrated algorithm based on machine learning techniques*”) јесте научно заснована, а резултати који ће проистећи израдом ове докторске дисертације представљаће значајан и оригиналан допринос међународној научној заједници. Тема припада ужој научној области Рачунарска техника и информатика.

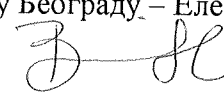
На основу свега претходно изложеног, Комисија са задовољством предлаже Наставно-научном већу Електротехничког факултета Универзитета у Београду да кандидату **Бранислави Цвијетић** одобри израду докторске тезе под насловом **Откривање структура података употребом интегрисаног алгоритма заснованог на техникама машинског учења** (на енглеском „*Discovering data structures using an integrated algorithm based on machine learning techniques*”), под менторством др Захарија Радивојевића, ванредног професора Електротехничког факултета Универзитета у Београду.

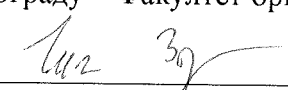
У Београду, 31.05.2024.

ЧЛАНОВИ КОМИСИЈЕ


др Захарије Радивојевић, ванредни професор
Универзитет у Београду – Електротехнички факултет


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


др Синиша Влајић, редовни професор
Универзитет у Београду – Факултет организационих наука


др Зоран Чича, редовни професор
Универзитет у Београду – Електротехнички факултет