

**Наставно-научном већу
Математичког факултета
Универзитета у Београду**

Одлуком Наставно-научног већа Математичког факултета Универзитета у Београду донетом на седници одржаној 22.09.2014. именовани смо у Комисију за преглед и оцену докторске дисертације "Информатички модели у анализи осећања засновани на језичким ресурсима" кандидата Миљане Младеновић, дипломираног инжењера електронике. После прегледања поднетог рукописа подносимо следећи

ИЗВЕШТАЈ

• Биографски подаци

Миљана Младеновић (рођ. Јосифов) рођена је 11. 12. 1963. године у Врању. Основну школу „Вук Караџић“ завршила је као ђак генерације. Гимназију „Бора Станковић“ у Врању завршила је 1982. као ђак генерације. Школске 1982/1983. године уписала је студије на Електронском факултету у Нишу (Одсек за рачунарску технику и информатику) и дипломирала 1986/1987. године са просечном оценом 9,02. Школске 2009/2010. године уписала је докторске студије на Математичком факултету, смер Рачунарство и информатика. Положила је све предмете предвиђене програмом студија са просечном оценом 10,00.

Радно искуство стицала је у компанијама: Симпо холдинг компанија у Врању на пословима програмер–приправник од децембра 1987. до новембра 1988, YUMCO холдинг компанија у Врању на пословима програмер–аналитичар од новембра 1988. до јануара 1992, Дуванска индустрија Врање на пословима Computer programmer од јануара 1992. до септембра 1992. Симпо Холдинг компанија у Врању, на пословима шефа рачунског центра од септембра 1992. до новембра 1998. Од новембра 1998. до данас је главни менаџер Агенције "Milenijum III", Врање. У периоду од новембра 1998. до септембра 2000. радила је и у Техничкој школи у Врању као професор рачунарства и информатике. Такође је радила и у организацијама "UNDP", Врање као Office Applications Trainer - од априла 2004. до новембра 2004, Development Alternatives Inc. USAID као Office Applications Trainer – од октобра 2005. до фебруара 2006. као и у компанији Guidance AD Beograd као Microsoft Certified Trainer for MCPD – од јуна 2007. до фебруара 2011. Додатна усавршавања стицана у току рада су: Microsoft Certified Professional Developer MCPD for Windows, Web and Enterprise applications, Achievement Date 2006. Certification Number: C775-9095, Microsoft Certified Trainer, Achievement Date 2008. Certification Number: D242-9741, Microsoft Certified Technology Specialist, Achievement Date 2009. Certification Number: C775-9102.

Основне области интересовања су јој обрада природних језика и њене примене, класификација текста, класификација текста према осећањима, семантички веб, XML-технологије, машинско учење и системи засновани на знању. Коаутор је дигиталног речника говора југа Србије објављеног 2013. године уз подршку Регистра националног Интернет домена Србије у оквиру пројекта 4ПИ.

• **Научни радови**

1. Antić, D., Mladenović, M., & Ivanov, S. (1997). Bondlab Graphical Editor - the Tool of the BONDLAB Software for Bond Graph Modelling. In *Proceedings of the 7th Symposium of Mathematics and its Applications*, 167–172. University of Timisoara. (M33)
2. Antić, D., Mladenović, M., & Stanković, L., (1997). Algorithmical assumptions of the dynamic systems modelling through the application of the BONDLAB software. In *Proceedings of the seventh Symposium of Mathematics and its Applications*, 13-18. University of Timisoara. (M33)
3. Antić, D., L. Stanković, B. Danković, M. Mladenović. (1997). Modeliranje i simulacija jedne klase hidrauličnih sistema primenom bond grafa. In *Proceedings of the XLI Conference ETRAN1*, 489-492. Zlatibor. (M63)
4. Antić, D., Vidojković, B., & Mladenović, M. (1999). An Introduction to Bond Graph Modelling of Dynamic Systems. In *Proceedings of TELSIKS '99, 4rd International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services (IEEE & MTT)*, 2, 661–664. (M33)
<http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=806297&url=http%3A%2F%2Fieeexplore.ieee.org%2Fiel5%2F6528%2F17421%2F00806297>
5. Mladenović, M., & Mitrović, J. (2013). Ontology of Rhetorical Figures for Serbian. *Text, Speech, and Dialogue, Volume 8082 of the series Lecture Notes in Computer Science*, 386–393. Springer ISSN 0302-9743, ISBN 978-3-642-40584-6, DOI 10.1007/978-3-642-40585-3_49. (M33) ISI/Web of Science
http://link.springer.com/chapter/10.1007/978-3-642-40585-3_49
6. Mladenović, M., & Mitrović, J. (2014). Semantic Networks for Serbian: New Functionalities of Developing and Maintaining a WordNet Tool. In G. Pavlović Lažetić, C. Krstev, I. Obradović & D. Vitas *Natural Language Processing for Serbian – Resources and Application*, 1-11. Matematički fakultet, Beograd. (M62)
<http://jerteh.rs/wp-content/uploads/2015/05/Mladenovic.pdf>
7. Mladenović, M., Mitrović, J., & Krstev, C. (2014). Developing and Maintaining a WordNet: Procedures and Tools. In H. Orav, C Fellbaum & P Vossan (Eds.), *Proceedings of Seventh Global WordNet Conference 2014*, 55–62. University of Tartu, Tartu, Estonia. ISBN 978–9949–32–492–7 (M33) ISI/Web of Science
http://gwc2014.ut.ee/proceedings_of_GWC_2014.pdf
8. Mladenović, M. Digital Dictionary of South Serbian Dialect. *Infoteka* 15 , 1 (September 2014), 2a-55a. ISSN: 1450-9687 (M52)
http://infoteka.bg.ac.rs/pdf/Eng/2014/INFOTHECA_XV_1_2014_42a-55a.pdf
9. Mladenović, M.(2016). Digital Dictionary of the South Serbian Dialect. *Infoteka, Journal for Digital Humanities*,1, ISSN: 1450-9687. Faculty of Philology, University of Belgrade.(M52)

<http://infoteka.bg.ac.rs/index.php/en/2014/no-1/253-digi-lni-r-cni-g-v-r-ug-srbi/613-digital-dictionary-of-the-south-serbian-dialect>

10. Mladenović, M., Mitrović, J., Krstev, C., & Vitas, D. (2015). Hybrid Sentiment Analysis Framework For A Morphologically Rich Language. *Journal of Intelligent Information Systems* (In Press, Available online 15 August 2015) ISSN 0925-9902, ISSN (online) 1573-7675, DOI 10.1007/s10844-015-0372-5. (M23) IF 2014 0.886.
<http://link.springer.com/article/10.1007/s10844-015-0372-5#>
11. Mitrović, J., Mladenović, M., & Krstev, C. (2015). Adding MWEs to Serbian Lexical Resources Using Crowdsourcing. *Poster at The 5th PARSEME general meeting*. Iași, Romania, 23-24 September 2015. (M34)
<http://typo.uni-konstanz.de/parseme/images/Meeting/2015-09-23-lasi-meeting/WG1-MITROVIC-MLADENOVIC-KRSTEV-poster.pdf>.
12. Mladenović, M., Mitrović, J., & Krstev, C. (2016). Introducing a Language-independent Model for Adding a New Semantic Relation Between Adjectives and Nouns in a WordNet. In *Proceedings of Eight Global WordNet Conference*, 218-225. ISBN 978-973-0-20728-6(M33)
<http://gwc2016.racai.ro/proceedings.html>
13. Mladenović, M.(2016). Ontological detection of rhetorical figures. *Infotheca, Journal for Digital Humanities*, ISSN: 1450-9687. Accepted for publication, 2016. (M52)

• Конференције

1. Душко Витас, Миљана Младеновић, Стана Ристић, Гордана Павловић-Лажетић, Милош Утвић. Дигитални документи у процесима лексикографске обраде, IX конференција Нове технологије и стандарди: дигитализација националне баштине, Београд, 2010.
2. Миљана Младеновић, Синиша Ненков, Момчило Златановић. еРечник - дигитални речник говора југа Србије, XII конференција Нове технологије и стандарди: дигитализација националне баштине, Београд, 2013.

• Предмет дисертације

Ова докторска дисертација, чији је основни задатак анализа осећања у тексту, представља истраживање које се односи на класификацију текстова на основу осећања (енг. sentiment classification) на нивоу докумената, на српском језику, вероватносном методом машинског учења мултиномијалне логистичке регресије, односно, максималне ентропије. Циљ овог истраживања је генерисање првог свеобухватног, флексибилног, модуларног система за анализу текстова на српском језику на основу осећања уз помоћ дигиталних ресурса као што су семантичке мреже, специјализовани лексикони и доменске онтологије. У том смислу истраживање је подељено у две фазе. У првој фази су развијане методе и програмски алати којима се открива поларитет осећања дословног значења текста. У овом делу рада се предлаже, примењује и оцењује нова метода редукције векторског простора предиктора који се користе у процесу класификације текста на основу осећања. Предложена метода редукције примењује се у класификационом моделу

максималне ентропије, а ослања се на употребу лексичко-семантичке мреже WordNet и специјализованог лексикона речи и израза који изражавају осећања. Предложена метода састоји се из два сукцесивна поступка. Први се односи се на проширење векторског простора предиктора флективним облицима предиктора. У истраживању је показано да у анализи осећања употреба стемера, као стандардног метода смањења векторског простора у класификаторима текста, може довести до непотпуних или нетачних ознака предиктора поларитета осећања, док се увођењем флективних облика предиктора то може избећи. У раду се, затим, показује да иницијално повећање векторског простора услед увођења флективних облика, може бити успешно редуковано применом другог предложеног поступка — семантичког пресликавања свих предиктора истог поларитета осећања у мањи број семантичких класа. Тиме се векторски простор предиктора редукује у односу на иницијални.

У другој фази дисертације описана је израда и имплементација формалних онтологија реторских фигура српског језика - доменске онтологије и онтологије задатака. Приказана је имплементација онтологије задатака у процесу генерисања предиктора фигуративног говора. Примењени су поступци закључивања дескриптивном логиком на основу правила дефинисаних у самој онтологији. Задатак истраживања друге фазе је препознавање фигуративног говора како би се унапредио постојећи скуп предиктора генерисан у првој фази истраживања. Резултати истраживања ове фазе показују да се неке класе стилских фигура могу аутоматски препознавати што може унапредити процес класификације на основу осећања.

У току рада на овој дисертацији пројектован је, имплементиран и статистички оцењен софтверски алат SAFOS (Sentiment Analysis Framework for Serbian) који представља интегрисани систем за класификацију текстова на основу осећања на српском језику. Евалуација модела извршена је унакрсном валидацијом и независним скуповима, а статистичка оцена t-тестом и z-тестом. Према нашим сазнањима, ово је први модел класификације на основу осећања текстова на српском језику, па је поређење извршено само са моделима добијеним истом методом на другим језицима. Резултати добијени унакрсном валидацијом бољи су или у рангу са моделима генерисаним у неколико последњих година на другим језицима. При примени независних скупова, добијени резултати такође су бољи или у рангу са резултатима који се добијају примењеном методом на другим језицима.

- **Приказ дисертације**

Докторска дисертација “Информатички модели у анализи осећања засновани на језичким ресурсима,” садржи седам поглавља, 25 прилога, 48 слика, 24 табеле и има следећу структуру. У **поглављу 1** уведени су и дефинисани појмови и методе који се користе у првој фази истраживања: класификација текста, класификација текста на основу осећања, машинско учење, надгледано машинско учење, вероватносни модели надгледаног машинског учења, модели репрезентације докумената. На крају уводног дела дефинисани су задаци и циљеви истраживања. У **поглављу 2** приказан је математички модел метода класификације текстова и класификације текстова на основу осећања. Дат је математички модел вероватносне класификације текста и примена вероватносне класификације у регресионим моделима. На крају поглавља приказан је математички модел метода максималне ентропије као једног од регресионих модела који се успешно примењује у задацима обраде природног језика. У **поглављу 3** приказани су лексички ресурси српског језика као и методе и алати за њихову обраду. **Поглавље 4** бави се свеобухватним истраживањем тренутно актуелних типова и метода класификације текстова на основу осећања. **Поглавље 5** приказује доприносе ове дисертације методама редукације векторског простора предиктора коришћених у класификационом моделу максималне ентропије. Најпре се врши анализа метода редукације предиктора. Предлаже се метода редукације којом се унапређује

класификација текстова на основу осећања. Дефинише се математички модел предложене методе. Уводе се и описују скупови за учење и скупови за тестирање. Дефинишу се лексичко-семантички ресурси који се користе у примени предложене методе. У поглављу 5 дат је опис изградње и евалуације система за класификацију текстова на основу осећања – SAFOS у коме се примењује и оцењује предложена метода редукције векторског простора предиктора. Дефинишу се параметари и функције система SAFOS-а, као и параметари за оцену система: прецизност, одзив, F1-мера и тачност. Даје се опис одабране методе за оцену статистичке значајности предложеног модела и начин примене над скуповима за тестирање у систему SAFOS. На крају поглавља дат је приказ изведених експеримената, резултата и оцена система. **Поглавље 6** односи се на поступке препознава фигуративног говора ради побољшања класификације текстова на основу осећања. Уведен је појам доменске онтологије и описан појам реторске фигуре. Размотрен је значај фигуративног говора у класификацији текстова на основу осећања. Дат је опис изградње и структуре прве формалне доменске онтологије реторичких фигура за српски језик RetFig.owl и одговарајуће онтологије задатака дефинисане на основу правила за препознавање одређених класа реторских фигура применом дескриптивне логике. На крају овог одељка дат је приказ изведених експеримената, резултата и оцена софтверског додатка систему SAFOS којим се препознаје фигуративни говор.

Закључно поглавље ове дисертације бави се приказом резултата, проблема и недостатака у систему SAFOS. Најзад, указује се и на технолошки, друштвени, педагошки и научни значај анализе осећања и препознавања фигуративног говора и дају даље смернице у процесу развоја система SAFOS.

Главни научни доприноси у приложеном раду су:

- ◆ Нова, хибридна метода редукције векторског простора предиктора примењена у бинарној класификацији текстова према осећањима;
- ◆ Пројектовање и имплементација првог модела класификације на основу осећања текстова на српском језику;
- ◆ Изградња прве формалне дескриптивне онтологије реторичких фигура у српском језику;
- ◆ Нова метода учења онтологије из дигиталног корпуса савременог српског језика;
- ◆ Нова метода аутоматског препознавања реторичких фигура у текстовима на српском језику помоћу онтологије Српски ворднет.

• Закључак

У рукопису “Информатички модели у анализи осећања засновани на језичким ресурсима,, кандидат Миљана Младеновић је показала систематично познавање области и проблема класификације података уопште и посебно класификације докумената према осећањима израженим на природном језику, као и креативну способност научног истраживача да своје теоријске резултате стави у функцију решавања актуелних информатичких проблема. Развијен је нови модел класификације текста према осећањима, предложен и статистички оцењен нови модел редукције векторског простора предиктора тог модела. Модел је, даље, унапређен препознавањем фигуративног говора уз помоћ доменске онтологије реторских фигура и онтологије Српски ворднет.

Кандидат је кроз овај рад дао теоријски, методолошки и практични допринос решавању

проблема класификације докумената на основу осећања, а тиме и широј области истраживања текста.

Предлажемо стога Наставно-научном већу Математичког факултета да рукопис “Информатички модели у анализи осећања засновани на језичким ресурсима,, прихвати као докторску дисертацију и одреди комисију за јавну одбрану.

У Београду, 26.02.2016.

др Душко Витас
ванредни професор Математичког факултета

др Гордана Павловић-Лажетић
редовни професор Математичког факултета

др Ненад Митић
ванредни професор Математичког факултета

др Владан Девеџић
редовни професор Факултета организационих наука

др Цветана Крстев
редовни професор Филолошког факултета