

НАСТАВНО-НАУЧНОМ ВЕЋУ ЕЛЕКТРОТЕХНИЧКОГ ФАКУЛТЕТА БЕОГРАД

Одлуком наставно-научног већа Електротехничког факултета у Београду, донетом 24.12.2021. године, одређени смо за чланове **Комисије за преглед и оцену докторске дисертације** **Мр Гума Абдулкахдер Лакшен**, под насловом „**Окружење за анализу и оцену квалитета великих и повезаних података**“. Након прегледа дисертације, подносимо Већу следећи

ИЗВЕШТАЈ

1. УВОД

1.1. Наслов и обим дисертације

Наслов дисертације је „*Окружење за анализу и оцену квалитета великих и повезаних података*“. Дисертација је написана на енглески језику (*A framework for analysis and quality assessment of a big and linked data*). Обим дисертације је 160 страна и садржи 34 слике, 20 табела и листу од 351 библиографске референце.

1.2. Хронологија одобравања и израде дисертације

Гума Абдулкахдер Лакшен је уписао докторске академске студије школске 2013/2014. године на Електротехничком факултету Универзитета у Београду, на модулу софтверски инжењеринг, где је положио све испите са оценом 9.50 и испунио све предвиђене обавезе планом и програмом докторских академских студија.

Кандидат Гума Абдулкахдер Лакшен је пријавио тему за израду докторске дисертације 02. маја 2019. године.

Комисија за студије трећег степена разматрала је предлог теме за израду докторске дисертације (07.05.2019. године) и предлог Комисије о оцени подобности теме и кандидата упутила Наставно –научном већу на усвајање. Наставно-научно веће именовало је Комисију за оцену услова и прихватање теме докторске дисертације (број одлуке 5054/13-1 од 23.05.2019. године).

Кандидат је положио јавну усмену одбрану теме (докторски испит) дана 11.06.2019. године.

Наставно-научно веће усвојило је Извештај Комисије за оцену услова и прихватање теме докторске дисертације (Одлука бр. 5054/13-2 од 09.07.2019. године).

Веће научних области техничких наука дало је сагласност на предлог теме докторске дисертације (26.08.2019. године, број одлуке 61206-2982/2-20 од 28.9.2020. године).

Кандидат је предао докторску дисертацију на преглед и оцену 02.12.2021. године.

Комисија за студије трећег степена потврдила је испуњеност потребних услова за подношење предлога Наставно-научном већу Електротехничког факултета за формирање Комисије за преглед и оцену докторске дисертације (07.12.2021. године).

Наставно-научно веће Факултета именовало је Комисију за преглед и оцену докторске дисертације (број одлуке 5054/13-3 од 24.12.2021. године) у саставу:

- Професор Др Сања Вранеш, редовни професор, Електротехнички факултет у Београду
- Др Бошко Николић, редовни професор, Електротехнички факултет у Београду
- Др Валентина Јанев, виши научни сарадник, Институт Михајло Пупин у Београду

На основу члана 101. Статута Универзитета у Београду, члана 74. Статута Универзитета у Београду-Електротехничког факултета и захтева студента, одобрено је продужење рока за завршетак студија до истека троструког броја школских година потребних за реализацију уписаног студијског програма.

1.3. Место дисертације у одговарајућој научној области

Дисертација је из области Рачунарске технике и информатике, за коју је матичан Електротехнички факултет Универзитета у Београду.

За ментора дисертације одређена је др Сања Вранеш, редовни професор Електротехничког факултета Универзитета у Београду. Ментор је одређен на основу остварених научних доприноса из области компјутерских наука и софтверског инжењерства, где тематски ова дисертација припада.

1.4. Биографски подаци о кандидату

Образовање:

Гума Абдулхадер Лакшен рођен је 1. јануара 1963. у месту Ал-Асаб , код Триполија у Либији. Основно и средњу школу је завршио од 1969 до 1982. У 1983. години Гума Абдулхадер Лакшен је добио стипендију за студије у Енглеској од стране новоосноване либијске компаније ЛИСКО за гвожђе и челик. Средњу школу завршава у Енглеској 1985. године са одличним резултатима посебно у математици. Студије продужава на одсеку за компјутерски инжењеринг на колеџу у Кардифу, на универзитету у Велсу, 1985-1989. У периоду од 2002 до 2006, Гума Абдулхадер Лакшен студира и завршава мастер студије на смеру за компјутерске системе на Академији Libyan Higher academy са оценом 3.79/4.00.

Стручно искуство:

Гума Абдулхадер Лакшен је своју каријеру започео радећи у LISCO-у од 1989. до 1997. године, током ових година углавном је радио као програмер, шеф компјутерске секције, и менаџер одељења за рачунарске системе (**ИКТ менаџер**, Libyan Iron and Steel

Company, Misurata, Libya). У периоду од 1998 до 2013. години радио је на следећим позицијама:

- **ИКТ Менаџер** (2000-2002) Al-Maya medical factory, GMPMSCO, Al-Maya, Libya
- **Менаџер** (2007-2009) Al-rabta Pharmaceutical factory, GMPMSCO Libya, Al-rabta, Libya
- **Менаџер логистике** (2009-2011) Libyan Pharmaceutical/medical Company, GMPMSCO, Tripoli, Libya
- **Лидер одбора за међународне односе и инвестиције** - GMPMSCO, где је сарађивао са компанијама из области медицине и фармације (Hemofarm-Serbia, International PharmaChem-Italy, Oman Medical Supplies and Services Co. LLC- Oman)
- **Аналитичар података** (2006-2011) Almadena media and information center, Ministry of information and telecommunications, Tripoli, Libya
- **Руководилац одељења / Професор** (2011-2013). Head of Computer section, Zintan teaching college, Al-Jabal Al-Gharbe University, Zentan, Libya

Аутор или коаутор је већег броја техничких решења насталих у оквиру комерцијалних пројекта, као и више радова из области докторске дисертације, од чега један (1) у међународном часопису са SCI листе и седам (7) у зборницима међународних конференција.

2. ОПИС ДИСЕРТАЦИЈЕ

2.1. Структура и садржај дисертације

Дисертација је подељена у шест поглавља, има насловну страну, садржај, увод, централни део и закључак. Текст рада је структурно подељен у шест поглавља, која су наведена редом:

1. Увод (енг. *Introduction*)
2. Семантички Веб и велики подаци (енг. *The Semantic Web Space and Big Data*)
3. Проблеми са квалитетом повезаних екосистима података (енг. *Quality Issues of Linked data Ecosystems*)
4. Развоја решења за консолидацију арапских повезаних скупова података о лековима (енг. *Towards Solution Development for "Consolidation of Arabic Linked Drug Datasets"*)
5. Резултати и анализа (енг. *Results and Analysis*) односно аналитичко дедуктивни и синтетичко-индуктивни прилаз закључивању.
6. Закључак (енг. *Conclusions and Future Directions*)

По својој форми и садржају, поднети рад задовољава све стандарде за докторску дисертацију.

2.2 Кратак приказ појединачних поглавља

У првом поглављу су представљени изазови везани за обраду великих (енг. *Big Data*) и отворених података (енг. *Open Data*), као и циљеви истраживања у контексту интеграције линкованих података (енг. *Linked Data*) посебно у софтверским апликацијама из фармацеутске индустрије.

У другом поглављу су најпре уведени основни појмови везани за семантички веб (енг. *Semantic Web*) и концепт екосистема великих и отворених података. Приказана је такође прегледна анализа алата и технологија за обраду великих података, као и предности и значај откривања знања из великих података. Како је циљ истраживања обogaћивање локалних података (из арапских земаља са информацијама са Веба), квалитет интерлинкованих података је од пресудног значаја.

У циљу дефинисања иновативне, флексибилне софтверске инфраструктуре за интеграцију отворених и приватних података једне фармацеутске компаније у облик семантичког језера података, у трећем поглављу разматрају се изазови обраде повезаних и великих података са становишта процене квалитета података. Као оригинални научни допринос, кандидат предлаже методологију и методске поступке за процену квалитета повезаних отворених података (енг. *Linked Open Data*) и идентификује потребне активности и поступке или процедуре за изградњу *Linked Data* софтверске апликације и имплементацију одговарајућих сервиса за процену квалитета. На крају, у овом поглављу представљене су методе које ће се користити за процену квалитета линкованих података са посебним фокусом на тачност/валидност, конзистентност и релевантност података повезаних применом отворених алата (енг. *open-source*).

На бази разматрања из претходног поглавља, у четвртом поглављу представљен је оригиналан приступ изградњи свеобухватног система за интеграцију и претраживање лекова. ALDDA (*Arabic Linked Data Drug Application*) интегрише информације из више извора у облику модуларне онтолошке базе знања. *Linked Data* апликација је имплементирана применом стандардних семантичких *Web* речника као што су *RDF*, *SPARQL*, *DBO*, као и семантичких модела из фармацеутске индустрије (*DrugBank*).

Пето поглавље сумира најзначајније резултате докторске тезе укључујући (1) архитектура софтверске апликације ALDDA; (2) процес консолидације података; (3) оцена квалитета линкованих података на бази интеграције са арапском *DBpedia*-јом.

У шестом поглављу дат је закључак у којем се даје сажет приказ основних доприноса докторске дисертације и правци будућег развоја.

Преглед коришћене литературе дат је на крају докторске тезе.

3. ОЦЕНА ДИСЕРТАЦИЈЕ

3.1 Савременост, оригиналност и значај

Напредак у домену семантичких технологија и *Linked Data* приступ је омогућио бројним великим компанијама као што су *Google*, *Facebook* и *Amazon* да имају способност да искористе информације генерисане интеграцијом података из више извора и остваре компетитивне предности на тржишту. Међутим, скупови који се користе у арапским земљама садрже податке који су кодификовани арапским писмом и њихова обрада захтева посебне приступе, укључујући и сервисе за проверу квалитета у сваком кораку рада са

подацима. Стога је предмет истраживања у овом раду анализа технологија за оцену квалитета линкованих података и њихова примена у фармацеутској индустрији.

Оригиналноћ доприноса тезе се односи на закључке у вези примене *Linked Data* технологија за потребе компанија који раде са отвореним подацима из арапских земаља. Примена семантичких технологија у развоју напредних система за управљање арапским скуповима података је недовољно испитано подручје, а акценат академског рада је углавном био усмерен на концептуални развој и унапређење сервиса за оцену квалитета интегрисаних података. Овакво истраживање је, наравно, употребиво и за шире потребе изван арапског подручја.

У докторској тези је разрађен методолошки оквир за развој *Linked Data* апликација узимајући у обзир потребе за консолидацијом података, специфичности скупова података и потребе фармацеутске индустрије. Такође, представљен је дизајн и прототип целокупне апликације (*Arabic Linked Data Drug Application*) и посебно верификација и валидација резултата са подацима из четири арапске земље (Ирак, Саудијска Арабија, Сирија и Либан).

3.2 Осврт на референтну и коришћену литературу

Кандидат је у току израде докторске дисертације опсежно анализирао релевантну литературу из области примене семантичких технологија за рад са великим и интерлинкованим подацима. Коришћена литература, чији је преглед дат у дисертацији, обухвата 351 библиографских јединица (стручне књиге, радове публиковане у домаћим и страним часописима, као и радове излагане на конференцијама). У оквиру наведене литературе, изложен је такође списак објављених радова кандидата Гума Абдулхадер Лакшен. Радови кандидата су објављени у међународним и националним часописима и изложени на међународним и националним конференцијама.

3.3 Анализа примењених научних метода и њихова адекватност за спроведено истраживање

У истраживањима у оквиру дисертације примењивано је више различитих метода:

1. Анализа доступне литературе за оцену квалитета линкованих података и отворених графова знања (нпр. *DBpedia*) односно аналитичко-дедуктивна метода;
2. Метода дескрипције за описивање функционалности сервиса за оцену квалитета у процесу трансформације и интеграције отворених и затворених (енг. *proprietary*) скупова података применом *Linked Data* приступа, укључујући и потребна објашњења синтетичког или аналитичког вида;
3. Метода класификације за разврставање анализираних софтверских алата за оцену квалитета података;
4. Системска анализа за дефинисање једног свеобухватног система за ефикасно управљање интерлинкованим подацима у фармацеутској индустрији, као и његових компоненти/елемената и процеса односно релевантно синтетичко (индуктивно) закључивање.

Практични приступ за имплементацију апликације базира на:

- коришћењем различитих метода за сваку фазу имплементације, укључујући визуелно моделовање (UML - *Unified Modeling*);
- методу брзог развоја (RAD - *Rapid Application Development*);
- примена технологија отвореног кода.

На основу анализе докторске дисертације, може се закључити да примењени општи и посебни научни методи одговарају спроведеном научно-истраживачком раду, као и теми и карактеру дисертације.

Кораци и методски поступци које је кандидат предузео у свом раду водили су, са једне стране, непосредном решавању идентификованих конкретних проблема, а са друге стране, водили су сагледавању и решавању проблема практичне реализације из фармацеутске индустрије, што дисертацији даје додатну тежину.

3.4 Оцена способности кандидата за самостални научни рад

На основу прегледане докторске дисертације Комисија процењује да је кандидат Гума Абдулхадер Лакшен у потпуности демонстрирао спремност и способност за самостални научноистраживачки рад, почевши од систематичног (системског) прегледа актуелне литературе, преко показане креативности у осмишљавању истраживачких метода, до детаљног тумачења добијених резултата.

Закључено је да резултати дисертације имају ширу примену у даљем развоју семантичких сервиса и њихову интеграцију са постојећим информационим системом компанија. Кандидат је препознао значај и актуелност теме истраживања и демонстрирао креативност и иновације у научном раду, што је потврђено бројним публикацијама на релевантним научним конференцијама и у научном часопису. На основу свега наведеног, Комисија верује да је кандидат показао висок и довољан степен способности да ради независне научне радове.

4. ОСТВАРЕНИ НАУЧНИ ДОПРИНОС

4.1 Приказ остварених научних доприноса

Доприноси ове дисертације се односе на област примене семантичких технологија у пословним информационим системима, с циљем побољшања квалитета сервиса за интеграцију информација из отворених скупова података. У оквиру докторске дисертације остварени су следећи научни доприноси:

1. **Развој методологије за процену квалитета повезаних података:** методологија је заснована на анализи функционалности постојећих алата и недостатака који су уочени у претходним случајевима практичне употребе. На основу прегледа литературе и резултата тестирања, предложена је методологија која се односи на целокупан процес обраде података о лековима, од екстракције података, преко аутоматског повезивања са позадинским знањем (базе *DrugBank* и *DBpedia*) до сервиса за претраживање и визуализацију резултата претраживања.
2. **Консолидовање (арапских) отворених скупова података о лековима:** Имплементиран је процес за интеграцију отворених података са базама података

компаније и њихово чување у облику семантичког језера података и графа знања (енг. *Knowledge Graph*). Даље, стандардни статистички индикатори о скуповима података имплементирани помоћу *SPARQL* упитног језика су доступни кориснику преко *Virtuoso SPARQL endpoint*. Процес екстракције, трансформације, чишћење и консолидације је имплементиран помоћу алата отвореног кода (*HTML Crawlers - JSOUP, XLS, OpenRefine*) и применом стандардних семантичких речника.

3. **ALDDA -QA Окружење:** У вези са проблемима квалитета отворених података, имплементиран је сервис за процену квалитета (енг. *Quality Assessment*) узимајући у обзир разноликост (*variety*), варијабилност (*variability*), истинитост (*veracity*), ваљаност (*validity*) и димензију вредност (*value*) повезаних података, концентришући се на питања квалитета података као што су тачност (*accuracy*), доследност (*consistency*) и релевантност (*relevancy*). Димензије квалитета (тачност, доследност и релевантност) су евалуаирани помоћу интеграције базе са арапском верзијом графа *DBpedia*.

4.2 Критичка анализа резултата истраживања

Свеобухватном анализом практичне примене *Linked Data* приступ за потребе фармацеутске индустрије која користи (арапске) скупове података, кандидат је открио да се приступ најчешће користе у решавању проблема интегрисања података унутар предузећа или између предузећа у умреженим (виртуалним) организацијама. Даље, семантичке технологије се користе за претраживање структурисаних, као и неструктурисаних података и за анотацију садржаја. Урађене анализе у вези квалитета отворених база знања, као што је арапска *DBpedia*, су значајне за даља истраживања у овој области и биле су објављене у међународном часопису са *SCI* листе *Computer Science and Information Systems*.

У докторској тези, кандидат је предложио софтверску архитектуру за интеграцију података о лековима, као и начине за естраховање додатних информација из отворених извора, што омогућава њихову даљу машинску анализу, са нагласком на побољшање квалитета линкованих података (рад *Arabic Linked Drug Dataset: Consolidating and Publishing*).

Имајући у виду предмет истраживања, полазне хипотезе и циљеве дисертације, као и добијене резултате, Комисија констатује да је кандидат успешно одговорио на постављене изазове и да резултати оправдавају почетна очекивања. Кандидат је такође уочио и предвидео могућности за даља истраживања. Сходно њиховој применљивости, доприноси дисертације се могу начелно поделити у две групе, оне који представљају унапређење ширих научних знања, и оне који су везани конкретно за напредак у решавању изазова у раду са подацима (посебно из арапских земаља).

4.3 Очекивана примена резултата у пракси

Предложени приступ за интеграцију, оцену квалитета линкованих података и претраживање података о лековима је оригиналан. Прототип система приказан у оквиру докторске тезе је значајан допринос посебно за кориснике из фармацеутске индустрије јер нуди ефикасан и квалитетан сервис за претраживање података о лековима из отворених извора (посебно из арапских земаља).

Крањи резултат, *семантичко језеро* података и граф знања (енг. *Knowledge Graph*) и ALDDA апликација проширују постојећи информациони систем предузећа и на бази обогаћивања постојећих ресурса са ресурсима са веба пружају додатне информације и бенефите у пословању. Резултати докторске тезе су валидирани са подацима из четири арапске земаље, док комплетно решење предвиђа укључивање и додатних скупова података.

4.4 Верификација научних доприноса

У директној вези са докторском дисертацијом кандидат је објавио осам радова.

Радови у часописима међународног значаја

- Guma Lakshen, Valentina Janev, and Sanja Vraneš. 2021. “Arabic Linked Drug Dataset: Consolidating and Publishing”. *Computer Science and Information Systems*. 2021. ComSIS Consortium. Volume 18, Issue 3, Pages: 729-748. <https://doi.org/10.2298/CSIS123456789X>

Радови саопштени на скупу међународног значаја, штампани у целини

- Guma Lakshen, Valentina Janev, and Sanja Vraneš. 2019. “Linking Open Drug Data: Lessons Learned”. *IFIP International Conference on Computer Information Systems and Industrial Management*. book: *Computer Information Systems and Industrial Management*. (pp.164-175). DOI:10.1007/978-3-030-28957-7_15
- Guma Lakshen, Valentina Janev, and Sanja Vraneš. “Linking Open Drug Data: The Arabic dataset”. 2019. “The Arabic dataset. In: Konjović, Z., Zdravković, M., Trajanović, M. (Eds.) *ICIST 2019 Proceedings*, pp.22-26.
- Guma Lakshen, Valentina Janev, and Sanja Vraneš. “Quality Issues of Open Big Data Ecosystems: Toward Solution Development”. 2019. “The Arabic dataset. In: Konjović, Z., Zdravković, M., Trajanović, M. (Eds.) *ICIST 2019 Proceedings*.
- Guma Abdulkhader Lakshen, Valentina Janev, and Sanja Vraneš. “Challenges in Quality Assessment of Arabic DBpedia”. 2018. *WIMS '18: Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics*. June 2018. Article-No.: 15. Pages 1–4. <https://doi.org/10.1145/3227609.3227675>
- Guma Lakshen, Valentina Janev, and Sanja Vraneš. 2016. “Big Data and Quality: A Literature Review”. *24th Telecommunications forum TELFOR 2016*. 22-23 Nov. 2016. IEEE. Belgrade, Serbia. DOI: 10.1109/TELFOR.2016.7818902.
- Basma Kajruba, Guma Lakshen, and Jelena Velickovic. 2021. “The impact of big data analysis in enhancing sustainable development goals”. *The 1st conference on sustainable development from an economical perspective*. Gharyan University, Libya. كلية الواقع المستدامة التنموية) حول الأول العلمي المؤتمر فعاليات تبدأ غريان جامعة المحاسبة | University of Gharyan (gu.edu.ly) – (اقتصادي منظور من والمأمول
- Mohiuddin, K., Mohammad, A.R., Sivarathri, S., Lackshe, J.A. 2013. “Digital intellectual property as a service (DIPaaS): For mobile cloud users”. *International Symposium on Computational and Business Intelligence, ISCBi 2013*, 2013, pp. 35–39, 6724319

5. ЗАКЉУЧАК И ПРЕДЛОГ

Докторска дисертација кандидата Гума Абдулхадер Лакшен је у целини написана у складу са образложењем наведеним у пријави теме и садржи све битне елементе који се захтевају Правилником о докторским студијама Електротехничког факултета Универзитета у Београду.

Текст дисертације је написан јасно и разумљиво, са добром организацијом поглавља. Циљеви дисертације прецизно су формулисани и дефинисани, а објављени резултати су потврдили испуњеност тих циљева, чиме је кандидат показао способност за самостални научни рад.

Комисија констатује да докторска дисертација кандидата Гума Абдулхадер Лакшен испуњава све законске, формалне и суштинске услове, као и све критеријуме који се примењују приликом вредновања докторске дисертације на Универзитету у Београду, и на Електротехничком факултету. Узимајући у обзир све остварене резултате и оригинални научни допринос, комисија предлаже Наставно-научном већу Електротехничког факултета Универзитета у Београду да се докторска дисертација Гума Абдулхадер Лакшен под насловом „*Окружење за анализу и оцену квалитета великих и повезаних података*“ изложи на увид јавности, прихвати и упути на коначно усвајање Већу научних области техничких наука Универзитета у Београду, а кандидату одобри јавна усмена одбрана.

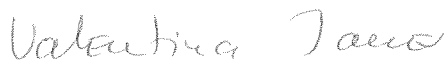
Београд, 29.12.2021.



Др Сања Вранеш, редовни професор
Универзитет у Београду - Електротехнички факултет



Др Бошко Николић, редовни професор
Универзитет у Београду - Електротехнички факултет



Др Валентина Јанев, виши научни сарадник
Институт Михајло Пупин