

УНИВЕРЗИТЕТ У БЕОГРАДУ
ФАКУЛТЕТ ОРГАНИЗАЦИОНИХ НАУКА

НАУЧНО-НАСТАВНОМ ВЕЋУ

Предмет: Реферат о урађеној докторској дисертацији кандидата Драгољуба Крнете

Одлуком бр. 3/106-4 од 05.11.2014. године, именовани смо за чланове Комисије за преглед, оцену и одбрану докторске дисертације кандидата Драгољуба Крнете под насловом

**Аутоматизација физичког пројектовања складишта података
проширеног Data Vault приступа**

После прегледа достављене Дисертације и других пратећих материјала и разговора са Кандидатом Драгољубом Крнетом, Комисија је сачинила следећи

РЕФЕРАТ

1. УВОД

1.1. Хронологија одобравања и израде дисертације

Одлука о усвајању извештаја Комисије о научној заснованости пријављене докторске дисертације донета је 11.07.2012., бр. 3/79-21. Одлуком Универзитета у Београду од 17.09.2012., бр. 06-20264/35-12 даје се сагласност на предлог теме докторске дисертације Драгољуба Крнете под називом „Аутоматизација физичког пројектовања складишта података проширеног Data Vault приступа“, а за ментора је именован проф. др Зоран Марјановић, редовни професор Факултета Организационих Наука, Универзитета у Београду. Ментор је известио Наставно научно веће Факултета Организационих Наука да је Драгољуб Крнета завршио израду

докторске дисертације, а Наставно научно веће Факултета Организационих Наука је именovalo Комисију за оцену завршене докторске дисертације 05.11.2014. бр. 3/106-4.

Кандидат Драгољуб Крнета је уписао докторске студије Факултету организационих наука Универзитета у Београду школске 2009/10. На докторским студијама положио свих девет испита са просечном оценом 9,78 предвиђених планом и програмом докторских студија на смеру Информациони системи. Приступни рад за докторску дисертацију је одбранио у јулу мјесецу 2012. године.

1.2. Научна област дисертације

Предмет дисертације је процес дизајнирања односно израде физичког модела складишта података. Централни проблем дисертације се односи на могућност аутоматизације процеса пројектовања физичког модела складишта података базираног на Data Vault концепту, а на основу модела метподатака извора података. Проблем истраживања се може експлицитно окарактерисати као метод аутоматизације физичке интеграције података, уз потребно очување историје промена.

Докторска дисертација припада области техничких наука и подручју организационих наука. Ужа научна област којом се бави дисертација је информациони системи и технологије. Ментор дисертације је проф. др Зоран Марјановић, редовни професор Факултета Организационих Наука, Универзитета у Београду.

1.3. Биографски подаци о кандидату

Драгољуб Крнета је рођен 06.01.1966. у селу Зебе општина Босанско Грахово, Босна и Херцеговина. Основну и средњу школу завршио је у Бос. Грахову као носилац дипломе „Огњен Прица“ (пандан дипломе „Вук Караџић“ у Србији). Завршио је Електротехнички факултет, смјер рачунари и аутоматика, први степен, на Универзитету у Бања Луци 1994. године. Технички факултет „Михајло Пупин“ у Зрењанину на Универзитету у Новом Саду завршава 2006. године, а на истом факултету 2008. године завршава постдипломске мастер студије са просечном оценом

9,23 и стиче звање Дипломирани инжењер пословне информатике – мастер. Мастер рад у области пословне интелигенције под менторством проф. др Драгице Радосав одбранио је са оценом 10. Аутор или коаутор је дванаест научних и стручних радова из области развоја софтвера, база података, data warehouse и система пословне интелигенције.

Програмирањем, развојем софтвера и базама података бави се од 1989. год. До 2002. год. аутор је двадесетак различитих софтверских пословних апликација које су имплементирани у више од 50 предузећа. Средином 2002. год. прелази у компанију Lanaco Информационе технологије Бања Лука на позицију Менаџера сектора за развој софтвера. За вријеме дванаестогодишњег рада у компанији Lanaco, активно је учествовао свим значајнијим софтверским пројектима било као пројект менаџер, пројектант информационог система и базе података, систем аналитичар или програмер. Поседује разне сертификате у информационам технологијама од којих су најзначајнији: Microsoft Certified Professional (MCP) и Microsoft Certified Database Administrator (MCDBA). Члан је асоцијација Institute of Electrical and Electronics Engineers (IEEE) и The Data Warehousing Institute (TDWI).

Подаци о запослењу:

- јули 2002. - *Lanaco* информационе технологије д.о.о.,
Бања Лука
Менаџер Сектора за развој софтвера
- април 2000. – јуни 2002. *Metropolis Media International*, Бања Лука
Пројектант информационог система, програмер,
администратор базе података и информационог
система
- јуни 1996. – март 2000. *Merx d.o.o.*, Бања Лука
Пројектант информационог система, програмер

Континуирана едукација:

- 2002: Едукација *Microsoft SQL Server 2000 Database Programming Course*
- 2003: Едукација *SQL Server 2000 Database Administration Course*
- 2004: Стечен сертификат - *Microsoft Certified Professional (MCP)*
- 2005: Сертификовани интерни аудитор за систем квалитета *ISO 9001*
- 2006: Учешће на конференцији *South East Europe Forum ICT* - Београд са објављеним радом
- 2007: Учешће на конференцији *Microsoft Worldwide Partner Conference 2007, Denver, USA*
- 2007: Учешће на конференцији *SYMOPIS 2007* – Златибор са објављеним радом
- 2007: Учешће на конференцији *South East Europe Forum ICT* – Херцег Нови са објављеним радом
- 2008: Стечен сертификат - *Microsoft Certified Database Administrator (MCDBA)*
- 2008: Учешће на *Conference on Information and Communication Technologies 2008* – Мостар са објављеним радом
- 2008: Учешће на конференцији *Infoteh 2008* – Јахорина са објављеним радом
- 2008: Учешће на *IEEE - 6th International Symposium on Intelligent Systems and Informatics (SISY 2008)* са објављеним радом
- 2009: Учешће на конференцији *Infoteh 2009* – Јахорина са објављеним радом
- 2009: Учешће на конференцији *ITEO* – Бања Лука са објављеним радом
- 2009: Учешће на *Regional Business Process and Knowledge Management Conference, Сарајево - 2009*
- 2010: Учешће на конференцији *Infoteh 2010* – Јахорина са објављеним радом
- 2010: Учешће на конференцији *Microsoft Worldwide Partner Conference 2010, Washington, USA*
- 2011: Учешће на конференцији *Infoteh 2011* – Јахорина са објављеним радом
- 2012: Едукација за сертификацију за управљање пројектима *Project Management Professional (PMP)*
- 2013: Едукација *Implementing a Data Warehouse with Microsoft SQL Server 2012*
- 2014: Едукација *Updating your Database management Skills to SQL Server 2014*
- 2014: Учешће на конференцији *Infifest 2014* – Будва

Учешће у пројектима

Кандидат Драгољуб Крнета је учествовао у више десетина софтверских пројеката као пројект менаџер, пројектант информационог система и базе података, систем аналитичар или програмер. У наставку су наведени најзначајнији пројекти у којима је кандидат учествовао:

Година или период	Пројекат	Наручилац	Позиција на пројекту
2002-2012	Lanaco ERP (v.1, v.2, v.3) Пословни информациони систем предузећа	Систем имплементиран у више од 200 предузећа	Пројект менаџер, Пројектант информационог система
2002-2004	Пословни информациони систем	Фонд здравственог осигурања Републике Српске	Пројект менаџер, Пројектант информационог система
2003- 2005	SOTAC (SOcial Sector Technical Assistance Project)	Влада Републике Српске Влада Федерације БиХ (систем имплементиран у 164 општине у БиХ)	Систем аналитичар, Пројектант информационог система
2005-2006	Пословни информациони систем	Интеграл Инжењеринг Бања Лука	Пројект менаџер, Пројектант информационог система и базе података
2006-2012	Пословни информациони систем (одржавање софтвера)	Фонд здравственог осигурања Републике Српске	Account manager



УНИВЕРЗИТЕТ У БЕОГРАДУ
ФАКУЛТЕТ ОРГАНИЗАЦИОНИХ НАУКА

2007	Пословни информациони систем	Електропренос БиХ	Пројект менаџер
2008	Информациони систем за вођење основних средстава	Генерални секретаријат Владе Републике Српске	Пројектант информационог система
2008	Софтвер за обрачун зарада	Рудник и Термолекетрана Гацко	Пројектант информационог система и базе података
2008-2009	Информациони систем за обрачун, наплату и тарифирање топлотне енергије	Топлана АД Бања Лука	Пројект менаџер
2009	Систем пословне интелигенције	Фонд здравственог осигурања Републике Српске	Пројект менаџер, Data warehouse пројектант
2010	Пословни информациони систем	Радио-телевизија Републике Српске	Пројектант информационог система
2011	Пословни информациони систем	Клинички центар Бања Лука	Пројектант информационог система
2011-2013	Технолошко унапређење пословног информационог система	Фонд здравственог осигурања Републике Српске	Пројект Менаџер, Пројектант информационог система и базе података
2014	Business Intelligence sistem	MojMarket & Tropic Banja Luka	Data warehouse konsultant

Осим тога, Драгољуб Крнета је 2014. је именован за члана комисије за измјену и допуну наставних планова и програма првог и другог циклуса студија Електротехничког факултета Универзитета у Бања Луци.

2. ОПИС ДИСЕРТАЦИЈЕ

2.1. Садржај дисертације

Докторска дисертација кандидата Драгољуба Крнете, под насловом „Аутоматизација физичког пројектовања складишта података проширеног Data Vault приступа“ написана је на 221 страни и структурирана у 7 поглавља, као и дела који се односи на коришћену литературу, прилоге, индекс слика (укупно 87 слика) и индекс табела (укупно 17 табела).

Садржај

1. УВОД	1
1.1. Предмет и проблем истраживања	1
1.2. Циљеви истраживања	7
1.3. Хипотезе у истраживању.	8
1.4. Методе истраживања.	9
1.5. Садржај дисертације.	10
2. ОСНОВНИ АСПЕКТИ СКЛАДИШТА ПОДАТАКА.	12
2.1. Мултидимензиони дизајн.	13
2.2. ETL процес.	19
2.3. Презентација података кориснику информација.	22
2.4. Пословна интелигенција и складиште података.	24
2.5. Архитектуре складишта података.	25
2.6. Методологије развоја складишта података.	33



УНИВЕРЗИТЕТ У БЕОГРАДУ
ФАКУЛТЕТ ОРГАНИЗАЦИОНИХ НАУКА

2.7. Просторна и временска складишта података.	37
2.8. Data Vault	38
2.8.1. Теоријска основа Data Vault модела.	45
2.8.2 Data Vault, Anchor моделовање и 6NF.. . . .	48
2.9. Квалитет складишта података.. . . .	51
2.10. Најзначајнији добављачи у области складишта података. . .	56
3. ПРЕГЛЕД ИСТРАЖИВАЊА У ОБЛАСТИ АУТОМАТИЗАЦИЈЕ ПРОЈЕКТОВАЊА СКЛАДИШТА ПОДАТАКА..	60
3.1. Преглед теоријских истраживања у области аутоматизације DW дизајна.	60
3.2. Преглед реализованих алата у области аутоматизације дизајна складишта података.	63
3.3. Критички осврт на досадашња истраживања у области аутоматизације дизајна складишта података	65
4. МОГУЋНОСТ АУТОМАТИЗАЦИЈЕ ФИЗИЧКОГ ПРОЈЕКТОВАЊА СКЛАДИШТА ПОДАТАКА (ДИРЕКТНИ ПРИСТУП ПРОЈЕКТОВАЊА СКЛАДИШТА ПОДАТАКА)	67
4.1. Приједлог оригиналног приступа за аутоматизацију пројектовања складишта података базираног на Data Vault концепту	71
4.2. Директни приступ пројектовања физичког модела складишта података	74
4.2.1. Концептуализација метаподатака.	77
4.2.2. Идентификација извора података.. . . .	82
4.2.3. Идентификација PDV типова.	86
4.2.4. Иницијално декларисање структуре PDV	89
4.3. Аутоматизација физичког дизајна дата мартова	94
4.3.1. Концептуализација Data Vault метаподатака	99



УНИВЕРЗИТЕТ У БЕОГРАДУ
ФАКУЛТЕТ ОРГАНИЗАЦИОНИХ НАУКА

4.3.2. Идентификација PDM (Physical Data Mart) типова	101
4.3.3. Декларисање data mart структуре	103
4.4. Аутоматизација ETL процеса	107
4.4.1. Аутоматизација ETL процеса: OLTP – DW.	112
4.4.1.1. Аутоматизација ETL процеса: OLTP – Data Vault	115
4.4.2. Аутоматизација ETL процеса: DW – data mart	117
4.4.2.1. Аутоматизација ETL процеса: Data Vault – data mart	121
5. ПРОТОТИП АПЛИКАЦИЈЕ ЗА ПОДРШКУ ПРОЈЕКТОВАЊА СКЛАДИШТА ПОДАТАКА	125
5.1. Ларманова метода развоја софтвера.	126
5.2. Релациони модел података	130
5.3. Трослојна архитектура апликације.	131
5.4. Десктоп апликација	132
5.5. Динамички SQL и генератор програмског кода	133
5.6. Развој апликације Direct PDV	134
5.6.1. Захтјеви	134
5.6.1.1. Вербални опис	135
5.6.1.2. Опис захтјева помоћу случајева коришћења.	136
5.6.2. Анализа	155
5.6.2.1. Резултати анализе дијаграма секвенци	156
5.6.2.2. Структура софтверског система.	158
5.6.3. Пројектовање	160
5.6.3.1. Пројектовање релационог складишта података	161
5.6.3.2. Пројектовање апликационе логике	167



УНИВЕРЗИТЕТ У БЕОГРАДУ
ФАКУЛТЕТ ОРГАНИЗАЦИОНИХ НАУКА

5.6.3.3. Пројектовање корисничког интерфејса.. . . .	167
5.6.4. Имплементација	186
5.6.5. Тестирање	188
6. ПРАКТИЧНА ПРОВЈЕРА РЕЗУЛТАТА ИСТРАЖИВАЊА	190
6.1. Захтјеви	190
6.2. Пројектовање складишта података	191
6.3. Пројектовање дата марта	197
6.4. Коришћење података у систему пословне интелигенције	202
6.5. Проширење складишта података са подацима из екстерних полуструктурираних извора података	206
6.6. Закључак везан за практичну провјеру резултата истраживања	210
7. ЗАКЉУЧНА РАЗМАТРАЊА.	212
7.1. Провјера тачности постављених хипотеза	213
7.2. Остварени научни доприноси	217
7.3. Могућност примјене рјешења и правци даљњих истраживања	220
8. ЛИТЕРАТУРА	222
Слике	235
Табеле	239
Додатак А. Садржаји табела RuleTypes и Rules	240
Додатак Б. Дијаграми секвенци	244
Биографија аутора	287
Изјава о ауторству	288
Изјава о истоветности штампане и електронске верзије докторског рада	289
Изјава о коришћењу	290

2.2. Кратак приказ појединачних поглавља

У **првом, уводном делу** дисертације дата је формулација проблема и предмета истраживања, мотивација за истраживање, описан је циљ истраживања и хипотезе у истраживању, те је укратко описан метод и фазе урађеног истраживања. У овом дијелу се наводе дефиниције основних појмова од интереса из шире области истраживања, а затим указује на ужу област истраживања (аутоматизација пројектовања складишта података) и изабрани проблем истраживања, при чијем решавању су остварени доприноси представљени овом дисертацијом.

У **другом делу** дисертације разматрају се основни аспекти складишта података. У овом делу се даје неколико дефиниција складишта података, садржај и циљеви изградње складишта података, те однос пословне интелигенције и складишта података и њихов значај за подршку одлучивању. Даље се разматрају основни елементи мултидимензионог дизајна, ETL процеса и начина презентација података кориснику информација. У другом делу је дат преглед и карактеристике неколико архитектура складишта података са нагласком на основни аспекти као и теоријску заснованост Data Vault концепта складиштења података. Осим тога, у овом делу је дат и кратак осврт на квалитет складишта података.

Најновија истраживања у области дизајна складишта података разматрају се у **трећем делу** дисертације. За проналажење литературе, у овом делу су кориштени најпознатији претраживачи научних радова са конференција и часописа као и сугестије пројектаната и истраживача који дуже раде у области складишта података ради комплетирања слике и експертне оцјене релевантности смјерница прије детаљног читања и личног истраживања. У овом делу рада се даје критички осврт на досадашње резултате истраживања, идентификује се и анализира проблем аутоматизације дизајна складишта података, те се даје предлог истраживања да би се дошло до решења.

У **четвртном, у централном делу рада**, се развија директан приступ пројектовања физичког модела складишта података базираног на Data Vault концепту, те се дефинише формализација, генерализација и стандардизација проширеног Data Vault приступа, при чему се за сваку врсту извора података стандардује модел и трансформације уз коришћење метамодела и одговарајућих метаподатака и правила.

У **петом делу** рада описан је начин реализације прототипа алата Direct PDV који служи за аутоматизацију пројектовања складишта података базираног на Data Vault концепту. Алат се реализује уз коришћење (у контексту индустријски веома релевантног) конкретне система за управљање базама података Microsoft SQL Server 2008 R2. Постоји неколико разлога за одабир овог система као што су: бројност имплементација, расположивост укључујући цену, те интензиван развој (сада већ верзија 2014). Технички одлична алтернатива Oracle била би од већег значаја само за највеће и финансијски најснажније кориснике.

У **шестом делу** се показује практична валидација приступа и провера резултата истраживања, те се разматра употребна вредност аутоматизације дизајна складишта података и практична провера реализованог алата Direct PDV у области здравственог осигурања.

Закључак, осврт на научне доприносе и правце даљих истраживања су дати у **седмом поглављу**. На крају рада је дата коришћена литература, те одговарајући прилози дисертације.

3. ОЦЕНА ДИСЕРТАЦИЈЕ

3.1. Савременост и оригиналност

Тема којом се бави докторска дисертација је савремена и веома популарна, како у научном тако и у практичном погледу. Истраживање, формализација и развој нове методологије за аутоматизацију пројектовања физичког модела складишта података има врло широк практичан значај који се огледа у могућности примене развијеног решења у практичном раду свих организација којима је потребан брз развој складишта података.

У раду је представљен нов алгоритам за аутоматизацију физичког пројектовања Data Vault-а заснован на моделу метаподатака и моделу правила за пројектовање на основу извора трансакционих података. Односи између ентитета у трансакционом систему и правила за развој складишта података базираног на Data Vault концепту су кључни за

аутоматизацију пројектовања физичког модела складишта података. Концептуализација метаподатака представљена је физичким моделом података који се уз мање измене може користити у пројектовању индивидуалних складишта података и представља основни методолошки допринос рада.

3.2. Осврт на референтну и коришћену литературу

У раду је коришћена обимна и релевантна домаћа и страна литература, те извори података са Интернета. Највећа пажња посвећена је публикацијама најрелевантнијих научних часописа у области информационих технологија, база података, складишта података и пословне интелигенције. Велики број коришћених извора публикован је на водећим конференцијама и у водећим часописима из наведених области (нарочито у часописима са SCI листе). Поред тога, коришћена је обимна литература коју су публиковали водећи светски издавачи (Springer, Wiley, Prentice-Hall, McGraw Hill, и др). У изради овог рада референцирано је 150 извора релевантних за обраду теме дисертације.

3.3. Опис и адекватност примењених научних метода

Основни метод истраживања приликом израде докторске дисертације био је сакупљање и проучавање доступне литературе, њена анализа и систематизација, а све то с циљем да се покаже оправданост и корисност развоја новог модела аутоматизације пројектовања складишта података. У изради дисертације од општих научних метода, коришћена је метода моделовања и аналитичко-дедуктивна метода. Метода моделовања је коришћена приликом израде модела метаподатака извора и складишта података, модела складишта података и практичне провере резултата истраживања. Код анализе података о постојећим решењима у области складишта података коришћена је аналитичко-дедуктивна метода. Поред наведених општих метода, у делу израде модела метаподатака извора података, коришћени су и посебни методолошки приступи као што су генерализација (апстракција у којој се скуп сличних типова објеката третира као генерички тип) и специјализација као инверзни поступак (у којем се за неки тип објекта дефинишу његови подтипови са специфичним особинама). Напримјер, скуп типова извора података по

структурираношћу извора (структурирани, полуструктурирани, неструктурирани) представљен је генеричким типом *DataStructureType*.

Резултати истраживања су презентовани текстуално, описивањем и приказани графички кроз дијаграме, табеле и слике са приказом упоредних резултата. Истраживање током израде докторске дисертације било је интердисциплинарног карактера, јер је укључивало више научних дисциплина и то методологију, информатику, теорију одлучивања, менаџмент и друге. Процес аутоматизације пројектовања складишта података разматран је са становишта методологије. Извршено је утврђивање могућности развоја система за аутоматизацију пројектовања складишта података за подршку одлучивању са становишта информатике и теорије одлучивања. На крају је коришћење развијеног прототипа о области здравственог осигурања, односно анализа потрошње лијекова, обухваћено менаџментом.

Начин на који је коришћена литература, њени извори и критички осврт јасно указују на адекватан истраживачки приступ. У реализацији софтверског решења за аутоматизацију пројектовања складишта података коришћена је објектно-оријентисана методологија пројектовања употребом UML језика за моделовање. Експериментална евалуација и примена реализованог решења је адекватно постављена и извршена ослањајући се на водеће стандарде за евалуацију модела.

На основу анализе докторске дисертације, може се закључити да примењене научне методе и технике одговарају, по свом значају и структури, теми дисертације и спроведеном истраживању.

3.4. Примењивост остварених резултата

Дисертацијом је демонстрирана евалуација приступа израдом прототипа апликације која је омогућила провјеру постављених циљева и хипотеза. Ова једноставна апликација, названа Direct PDV, омогућује веома брзо пројектовање складишта података укључујући и пројектовање data marta и ETL процеса. На овај начин добијамо интеграцију свих потребних активности које су неопходне за реализацију комплетног складишта података. Осим тога, показано је проширење Data Vault приступа из извора података који нису базе података, односно из полуструктурираних извора података.

Демонстриран је приступ у пракси кроз коришћење прототипа апликације на реалном примјеру из области здравственог осигурања.

Коришћењем прототипа апликације показано је веома брзо и ефикасно пројектовање складишта података, data marta и ETL процеса. У овом делу је приказан реалан пример из здравственог осигурања, уз додатну анализу OLAP алатом и пример доношење пословних одлука на основу добијених информација. Практичном провером резултата истраживања показане су све предности аутоматизације пројектовања складишта података. Осим тога, алат се може користити и за развој прототипова складишта података где корисник може да брзо препозна радно окружење за анализу информација и самим тим постаје више заинтересован за развој складишта података. Пројектанти складишта података на тај начин могу да лакше откривали потребе корисника за информацијама.

3.5. Оцена достигнутих способности кандидата за самостални научни рад

Кандидат Драгољуб Крнета је у периоду од 2006. до уписа на докторске студије крајем 2009. године објавио девет стручних или научних радова на домаћим и међународним ИТ конференцијама.

Кандидат је на докторским студијама показао способност да методолошки адекватно конципира истраживање, критички анализира референце и резултате истраживања и на основу тога долази до закључака значајних за дисертацију. Током докторских студија објавио је као коаутор следеће радове:

- Крнета Д., Опачић В., Бобанац З., Бановић Н., Систем пословне интелигенције у Фонду здравственог осигурања, Infoteh 2010.
- Крнета Д., Марјановић З., Могућност интеграције методолошких приступа развоја информационог система и имплементације готовог ERP рјешења, Infoteh, 2011.

Из области која је предмет докторске дисертације написао је рад који је објављен у часопису са SCI листе:

Драгољуб Крнета, Владан Јовановић, Зоран Марјановић "A Direct Approach to Physical Data Vault Design", Computer Science and Information Systems (ComSIS), Vol. 11, No. 2, 569–599, (2014). ISSN: 1820-0214. IF (2013): 0.575

Као Менаџер сектора за развој софтвера компаније Lapaso IT Бања Лука учествовао је на многим ИТ скуповима од који су најзначајнији *Microsoft Worldwide Partner Conference* у Денверу 2007. и у Вашингтону 2010. године.

Као што се из приложене биографије може видети, кандидат Драгољуб Крнета је од 2006. године активан учесник у научно-истраживачком раду. Током стручне каријере, највећи део радних активности усмерио је на научно-истраживачки рад из области развоја информационих система, складишта података и система пословне интелигенције с обзиром да је запослен у компанији која се бави развојем и имплементацијом информационих система. Узевши у обзир целокупно залагање и остварене резултате у научно истраживачком раду, закључујемо да је кандидат способан да се у потпуности самостално бави научно-истраживачким радом.

4. ОСТВАРЕНИ НАУЧНИ ДОПРИНОСИ

4.1. Приказ остварених научних доприноса

Концептуализација метаподатака представљена физичким моделом података који се уз мање измене може користити у пројектовању индивидуалних складишта података, представља основни **методолошки допринос рада**. Приступ је могуће применити на различите типове организација и различите системе за управљање базама података.

Кроз истраживања, рад и резултате докторске дисертације остварени су следећи научни доприноси:

Преглед истраживања у области аутоматизације складишта података (поглавље 3), поред прегледа теоријских истраживања, даје се преглед реализованих алата у области аутоматизације дизајна складишта података. Осим тога, у поглављу 2 дати су основни аспекти складишта података везани за мултидимензиони дизајн, архитектуре, методологије развоја, те однос пословне интелигенције и складишта података. Овај део може да се користи као сажет увод у области складиштења података.

Формалан опис модела и метода који омогућују што већу аутоматизацију израде физичког модела складишта података базираног на Data Vault концепту представља централни део рада који је дат у поглављу 4 као директни приступ пројектовања складишта података. Приступ је дефинисан и образложен одговарајућим алгоритмом. Најпре је одређен проблем (ручно пројектовање складишта података), а затим је презентовано идејно решење кроз дијаграме, алгоритме, моделе и одговарајућа правила како би се аутоматизовао процес пројектовања складишта података.

Формализација проширења Data Vaulta у циљу принципијелног обухвата извора података из извора који нису само релационе базе података подразумевала је укључивање полуструктурираних извора података. Формализацију олакшава чињеница да, без обзира у колико извора се налазе оперативни подаци, сваки пословни систем тежи да интегрише, барем концептуално, основне податке од ширег значаја и успостави контролу, односно да има једну примарну базу података у којој је смештен највећи део података (а та база је у највећем броју случајева релациона). Полуструктурирани подаци се читавају у прелазну базу података, а затим у складиште података. Неструктурирани подаци морају проћи процесе текстуалне анализе, а затим се након структурирања пребацују у прелазну базу података. Даљњи поступак је сличан као код полуструктурираних података.

Унапређењу процеса дизајнирања складишта података. Аутоматизацијом дизајнирања складишта података знатно се убрзава развој комплетног система који омогућује израду прототипа складишта података у раној фази контакта са корисницима. На тај начин корисници постају више заинтересовани за давање информација. Пројектанти складишта података на тај начин могу лакше откривати знања о пословним процесима, те потребе корисника за информацијама.

Целокупан интегрисан приступ аутоматизацији пројектовања складишта података. Поред аутоматизације пројектовања самог складишта података, приступ омогућује пројектовање data marta и пројектовање ETL процеса (из извора у складиште података, те из складишта података у data mart). На овај начин добијамо интеграцију свих потребних активности које су потребне за реализацију комплетног складишта података.

Подизање квалитета пројектовања DW система, смањење трошкова и времена имплементације DW система. Квалитет пројектовања DW система показан је коришћењем GQM парадигме у процесу пројектовања складишта података. Примјеном GQM парадигме могу се дефинисати основни циљеви побољшања процеса пројектовања, развоја или одржавања складишта података, питања која се постављају за испуњење тих циљева, те предложена метрика. Аутоматским генерисањем програмског кода смањује се број потребних ИТ стручњака за реализацију система, а тиме се смањују вријеме и трошкови имплементације система.

Кроз истраживања, рад и резултате докторске дисертације остварени су и следећи стручни доприноси:

Демонстрирана је евалуација приступа израдом прототипа апликације која је омогућила проверу постављених циљева и хипотеза. Ова једноставна апликација, названа *Direct PDV*, омогућује веома брзо пројектовање складишта података укључујући и пројектовање data marta и ETL процеса.

Демонстриран је приступ у пракси кроз коришћење прототипа апликације на реалном примеру из области здравственог осигурања. Коришћењем прототипа апликације показано је веома брзо и ефикасно пројектовање складишта података, data marta и ETL процеса. У овом делу је приказан реалан пример из здравственог осигурања, уз додатну анализу OLAP алатом и пример доношење пословних одлука на основу добијених информација.

4.2. Критичка анализа резултата истраживања

Израда ове докторске дисертације подразумевала је обраду комплексне проблематике везане за аутоматизацију пројектовања складишта података базираног на **Data Vault** концепту. Кандидат је анализирао и систематизовао сазнања истраживањем обимне литературе и реализовао веома комплексно истраживање везано за могућности аутоматизацију пројектовања складишта података.

Дисертација је урађена у складу са пријавом, при чему је реализован циљ истраживања и потврђене хипотезе, главна и појединачне. Дисертација је

результат самосталног истраживачког рада кандидата. Примена утврђеног алгоритма за аутоматизацију пројектовања складишта података које је заснована на моделу метаподатака извора података проверена је реализацијом одговарајућег алата за пројектовање складишта података, а успешност истог је проверена емпиријски применом у области здравственог осигурања.

Прототип апликације развијен у овом раду служи за евалуацију предложеног приступа директног пројектовања складишта података и показује да приступ изводљив у пракси. Код коришћења прототипа апликације веома је битно да пројектанти складишта података пажљиво изврше избор домена за који се пројектује складиште података. Складиште података које је засновано на Data Vault концепту, највише има смисла користити у случају више дистрибуираних извора података. Кључна побољшања која су могућа на прототипу апликације односе се пре свега на реализацију функционалности која се односе на обухват сложенијих полуструктурираних и неструктурираних извора података, те реализацију функционалности текстуалне анализе неструктурираних података. С обзиром да је прототип апликације предвиђен само за рад на Windows оперативном систему, постоји простор за унапређење или развој потпуно нове апликације, која ће се ослањати на принципе, алгоритме и случајеве коришћења представљене овим радом, а која може да ради и на другим оперативним системима.

4.3. Верификација научних доприноса

Области научног интересовања Драгољуба Крнете су складишта података, односно аутоматизација пројектовања складишта података. Осим тога, кандидат има изузетна интересовања у области развоја информационих система и система пословне интелигенције.

Током свог научно истраживачког рада објавио је као аутор или коаутор дванаест стручних или научних радова на домаћим и међународним конференцијама и научним часописима од којих је најзначајнији рад везан за тему дисертације из категорије M23: Крнета Д., Јовановић В., Марјановић З., „*A Direct Approach to Physical Data Vault Design*“, објављен у часопису *Computer Science and Information Systems (ComSIS)*, Vol. 11, No. 2, 569–599, (2014). са *impact factorom* (IF) 0.575 за 2013. год.

5. ЗАКЉУЧАК И ПРЕДЛОГ

5.1. Кратак осврт на дисертацију у целини, на научне доприносе конкретној научној области и методологији, на примену у пракси и на способности кандидата

Докторска дисертација се бави значајном и још увек недовољно истраженом области аутоматизације пројектовања складишта података, а посебно базираном на Data Vault концепту, из структурираних и једноставнијих полуструктурираних и неструктурираних извора података.

Упоредним прегледом постојећих приступа у обрађеној области дата је тренутна слика академских активности у овој области. На основу анализе објављених истраживања идентификован је простор за предлог нове методологије за унапређење процеса пројектовања складишта података.

Основна идеја дисертације је била да се процес пројектовања складишта података формализује, генерализује и у значајној мјери аутоматизује. Приступ је дефинисан и образложен одговарајућим алгоритмом. Најпре је одређен проблем (ручно пројектовање складишта података), а затим је представљено идејно решење кроз дијаграме, алгоритме, моделе и одговарајућа правила како би се аутоматизовао процес пројектовања складишта података.

Полазна хипотеза аутора је била: *„Процес пројектовања складишта података базираног на Data Vault концепту и израде димензионог модела складишта података може се формализовати, генерализовати и у одређеној мјери аутоматизовати на основу физичког модела структурираних, полуструктурираних и једноставнијих неструктурираних извора података, укључујући и трансакционе базе података.*

Ради доказивања хипотезе развијени су алгоритми подржани концептуализацијом метаподатака која је представљена физичким моделом података који се уз мање измене може користити у пројектовању индивидуалних складишта података, што представља основни методолошки допринос рада.

Формалан опис модела и метода који омогућују што већу аутоматизацију израде физичког модела складишта података базираног на Data Vault концепту представља централни дио рада који је дат у поглављу 4 као

директни приступ пројектовања складишта података. У раду је представљена и формализација проширења Data Vaulta у циљу принципијелног обухвата извора података из извора који нису само релационе базе података подразумејевала је укључивање полуструктурираних извора података. Формализацију је олакшала чињеница да, без обзира у колико извора се налазе оперативни подаци, сваки пословни систем тежи да интегрише, барем концептулно, основне податке од ширег значаја и успостави контролу, односно да има једну примарну базу података у којој је смјештен највећи дио података (а та база је у највећем броју случајева релациона). Полуструктурирани подаци се учитавају у прелазну базу података, а затим у складиште података.

Неструктурирани подаци морају проћи процесе текстуалне анализе, а затим након структурирања пребацују у прелазну базу података. Даљњи поступак је сличан као код полуструктурираних података.

Поред аутоматизације пројектовања самог складишта података, приступ омогућује пројектовање data marta и пројектовање ETL процеса (из извора у складиште података, те из складишта података у data mart). На овај начин добијамо интеграцију свих потребних активности које су потребне за реализацију комплетног складишта података. Аутоматизацијом дизајнирања складишта података знатно се убрзава развој комплетног система који омогућује израду прототипа складишта података у раној фази контакта са корисницима. На тај начин корисници постају више заинтересовани за давање информација. Пројектанти складишта података на тај начин могу лакше откривати знања о пословним процесима, те о потребе корисника за информацијама.

Истовремено, у докторској дисертацији посвећена је пажња квалитету пројектовања и квалитету самог складишта података. Квалитет пројектовања DW система показан је коришћењем GQM парадигме у процесу пројектовања складишта података. Применом GQM парадигме могу се дефинисати основни циљеви побољшања процеса пројектовања, развоја или одржавања складишта података, питања која се постављају за испуњење тих циљева, те предложена метрика.

Аутоматско генерисање кода за креирање табела и кода за ETL процесе, смањује се број потребних ИТ стручњака, а тиме се смањују време и трошкови имплементације складишта података.

Примена утврђених алгоритама за аутоматизацију пројектовања складишта података који су засновани на моделу метаподатака извора података проверена реализацијом одговарајућег алата за пројектовање складишта података, а успешност истог је проверена емпиријски применом у области здравственог осигурања уз коришћење структурираних и полуструктурираних извора података.

С обзиром на постигнуте резултате, комплексност, актуелност и специфичност обрађене теме, ова дисертација задовољава највише критеријуме и показује способност кандидата Драгољуба Крнете за научно-истраживачки рад.

5.2. Предлог Комисије Научно-наставном већу

Предлажемо да се приложена докторска дисертација под називом „Аутоматизација физичког пројектовања складишта података проширеног Data Vault приступа“ кандидата Драгољуба Крнете прихвати, изложи на увид у јавности и упути на коначно усвајање већу техничких наука Универзитета у Београду.

ЧЛАНОВИ КОМИСИЈЕ

Проф. др Зоран Марјановић, редовни професор
Универзитет у Београду, Факултет организационих наука

Проф. др Милија Сукновић, редовни професор
Универзитет у Београду, Факултет организационих наука

Проф. др Владан Јовановић, редовни професор
Georgia Southern University, Allen E. Paulson College of Engineering and
Information Technology, USA