

28.06.2013. године
Београд,

УНИВЕРЗИТЕТ У БЕОГРАДУ
Веће научних области техничких наука

Београд
Студентски трг бр 1

У прилогу дописа достављамо потребну документацију за давање сагласности на предлог теме докторске дисертације кандидата Марка Петровића.

Захтев се доставља у складу са чланом 54 Закона о Универзитету.

С поштовањем,

СЕКРЕТАР ФАКУЛТЕТА

Гордана Алексић Славински, дипл. правник

Факултет организационих наука

УНИВЕРЗИТЕТ У БЕОГРАДУ

04-03-11/
(Број захтева)

Веће научних области техничких наука

(Назив стручног већа коме се захтев упућује сходночл 6. и
члану 7. ст. 1. овог Правилника)

28.06.2013.
(Датум)

ЗАХТЕВ
за давање сагласности на предлог теме докторске дисертације

Молимо да, сходо члану 57. ст. 3. Закона о универзитету («Службени гласник РС» бр. 21/02), дате сагласност на предлог теме докторске дисертације: (пун назив предложене теме докторске дисертације)
РАЗВОЈ ПРОЦЕСА ЕКСТРАКЦИЈЕ, ТРАНСФОРМАЦИЈЕ И ПУЊЕЊА ПОДАТАКА СКЛАДИШТА ПОДАТАКА ЗАСНОВАНА НА МОДЕЛОМ ВОЂЕНОМ ПРИСТУПУ

НАУЧНА ОБЛАСТ: ОРГАНИЗАЦИОНЕ НАУКЕ

ПОДАЦИ О КАНДИДАТУ:

1. Име, име једног од родитеља и презиме кандидата:
Марко (Властимир) Петровић
2. Назив и седиште Факултета на коме је стекао високо образовање: ФОН
3. Година дипломирања: 2002. година
4. Назив приступног рада кандидата: Развој ETL процеса заснован на MDD приступу
5. Назив факултета на коме је приступни рад одбрањен: ФОН
6. Година одбране приступног рада: 2013. година
7. Назив факултета на коме је кандидат завршио докторске студије: /
Одсек, смер или група: /
Година завршетка докторских студија: _____

Обавештавамо вас да је: НАСТАВНО НАУЧНО ВЕЋЕ ФОН-А
(назив надлежног тела факултета)

на седници одржаној: 28.06.2013. године размотрило предложену тему и закључило да је тема подобна за израду докторске дисертације.

ДЕКАН ФАКУЛТЕТА

Проф. др Милан Мартић

Прилог:

1. Предлог теме докторске дисертације са образложењем.
2. Акт надлежног тела факултета о подобности теме за израду докторске дисертације.

Поводом предлога Већа студијског програма докторских академских студија, а на основу члана 56. Статута Факултета, на седници Наставно-научног већа одржаној 28.06.2013. године, једногласно је донета следећа

О Д Л У К А

Усваја се Извештај Комисије за оцену научне заснованости пријављене докторске дисертације кандидата на докторским студијама Марка Петровића под насловом »РАЗВОЈ ПРОЦЕСА ЕКСТРАКЦИЈЕ, ТРАНСФОРМАЦИЈЕ И ПУЊЕЊА ПОДАТАКА СКЛАДИШТА ПОДАТАКА ЗАСНОВАНА НА МОДЕЛОМ ВОЂЕНОМ ПРИСТУПУ«. За ментора се предлаже др Зоран Маријановић, ред. проф. ФОН-а. Извештај се доставља Већу научних области техничких наука Универзитета у Београду.

Председник Наставно научног већа

Проф. др Милан Мартић

Доставити:

- Стручном сараднику за ПДС
- Ментору
- Вишем референту органа управљања и стручних органа
- Архиви



Број индекса: 14/2006

04-03-11/113

Датум: 19.06.2013.

На основу члана 161. Закона о општем управном поступку и службене евиденције, издаје се

УВЕРЕЊЕ О ПОЛОЖЕНИМ ИСПИТИМА

Марко Петровић, име једног родитеља Властимир, рођен 07.08.1977.године, Београд, Београд-Савски Венац, Србија, 2006/07. школске године уписан на докторске академске студије на 1. годину, 2012/13. школске године први пут уписан у 3. годину, самофинансирање, студијски програм Организационе науке, Информациони системи, током студија положио је испите из следећих предмета:

Р.бр.	Шифра	Назив предмета	Оцена	ЕСПБ	Фонд часова**	Датум
1.	ДС06МПРГ	Методологија програмирања	10 (десет)	10	II:(45+0+60)	20.02.2003.*
2.	ДС06РЗПЈ	Развој програма и програмски језици	10 (десет)	10	II:(45+0+60)	10.03.2003.*
3.	ДС06ИПОИ	Изабрана поглавља из операционих истраживања	10 (десет)	10	II:(45+0+60)	14.04.2004.*
4.	ДС06УПМЕ	Управљање предузећем - менаџмент	10 (десет)	10	II:(45+0+60)	15.07.2004.*
5.	ДС06ПРОС	Пројектовање организационе структуре	10 (десет)	10	III:(45+0+60)	01.02.2011.
6.	ДС07037	Статистика у менаџменту	10 (десет)	10	I:(45+0+60)	22.02.2011.
7.	ДС06МОНИ	Методологија и организација научних истраживања	10 (десет)	10	III:(45+0+60)	08.07.2011.
8.	Д00002	Развој информационих система	10 (десет)	10	III:(45+0+60)	29.11.2012.
9.	ДС07001	Управљање подацима	10 (десет)	10	I:(45+0+60)	05.04.2013.

* - еквивалентiran/признат испит.

** - фонд часова је у формату (предавања+вежбе+остало).

Општи успех: 10,00 (десет и 00/100) , по годинама студија (10,00, 10,00, /).

Виши стручни сарадник за докторске студије

др Марина Јовановић-Миленковић

Предмет: Подобност теме и кандидата Марка Петровића за израду докторске дисертације

Одлуком 3/57-12 бр. од 17.04.2013. године, именовани смо за чланове Комисије за оцену подобности теме и кандидата **Марка Петровића** за израду докторске дисертације и научне заснованости теме **„РАЗВОЈ ПРОЦЕСА ЕКСТРАКЦИЈЕ, ТРАНСФОРМАЦИЈЕ И ПУЊЕЊА ПОДАТАКА СКЛАДИШТА ПОДАТАКА ЗАСНОВАН НА МОДЕЛОМ ВОЂЕНОМ ПРИСТУПУ“**.

На основу материјала приложеног уз Захтев кандидата, Комисија подноси следећи

ИЗВЕШТАЈ

1. Подаци о кандидату

Марко Петровић је рођен 07. августа 1977. године у Београду. Завршио је Трећу београдску гимназију („Свети Сава“) у Београду, природно-математички смер. Након завршене гимназије, 1996. године уписао је Факултет организационих наука, смер информациони системи. Дипломирао је 2002. године са просечном оценом 8,19 а дипломски рад под насловом „Развој Веб Сервиса коришћењем .NET платформе“ је одбранио са оценом десет. Последипломске (магистарске) студије на Факултету организационих наука, смер информациони системи уписао је 2002. године, а затим 2006. године уписује докторске студије на Факултету организационих наука, смер информациони системи. Положио је све испите предвиђене планом и програмом, са просечном оценом 10.

Радно искуство

- Од 2001. године ангажован је као сарадник на Факултету организационих наука
- Од 2003. године запослен на Факултету организационих наука у звању асистент-приправник
- Од 2007. године запослен Факултету организационих наука у звању асистент

Наставне активности

Од 2000. године до 2003. године кандидат је радио као сарадник на Факултету организационих наука у Београду, на предметима Принципи програмирања и Пројектовање програма. Од 2003. године до 2007. године је радио као асистент-приправник на Факултету организационих наука у Београду, на предметима Принципи програмирања, Увод у информационе системе, Архитектура рачунара и оперативни системи, Програмски језици и преводиоци и Пројектовање програма. Од 2007. године до данас ради као асистент на Факултету организационих наука у Београду на предметима Програмски језици и Основе информационо комуникационих технологија.

Приликом евалуације од стране студената његов педагошки рад је редовно оцењиван високом оценом.

Приказ наставних активности у оквиру докторских студија

Кандидат Марко Петровић је школске 2006/2007 уписао докторске студије на Факултету Организационих наука – смер информациони системи. Положио је све планом и програмом предвиђене испите са просечном оценом 10 и то:

• Методологија програмирања	10 ЕСПБ
• Развој програма и програмски језици	10 ЕСПБ
• Изабрана поглавља из операционих истраживања	10 ЕСПБ
• Управљање предузећем – менаџмент	10 ЕСПБ
• Пројектовање организационе структуре	10 ЕСПБ
• Статистика у менаџменту	10 ЕСПБ
• Методологија и организација научних истраживања	10 ЕСПБ
• Развој информационих система	10 ЕСПБ
• Управљање подацима	10 ЕСПБ

Кандидат је одбранио приступни рад под називом „Развој ЕТЛ процеса заснован на МДД приступу“ и остварио 30 ЕСПБ бодова. Укупно је остварио 120 ЕСПБ бодова са просечном оценом 10.

Преглед публикованих и објављених радова

Марко Петровић објавио је, у сарадњи са другим ауторима, више научних радова у часописима међународног значаја, као и у зборницима са домаћих и међународних конференција.

Рад објављен у часопису међународног значаја (SCIE листа)

- М. Вучковић, М. Петровић, Н. Турајлић, М. Станојевић, „**The Specification of ETL Transformation Operations based on Weaving Models**“, Int J Comput Commun, ISSN 1841-9836, 7(5): 968-975, 2012. IF 2011 – 0.438 (M23)

Рад објављен на међународној конференцији

- Н. Турајлић, М. Петровић, М. Вучковић, И. Драговић, „**Groundwork for Presentation Pattern Metamodels**“, Зборник радова InfoTeh, Јахорина, 2013. (прихваћен рад)

Рад објављен у домаћем часопису

- М. Петровић, Н. Турајлић, И. Драговић, „**Преглед и упоредна анализа презентационих патерна**“, часопис ИнфоМ 9(34), 35-41, Београд, 2010.
- С. Нешковић, М. Петровић, „**Моделовање пословних процеса коришћењем OMG BPMN 2.0 стандарда**“, часопис ИнфоМ 8(31), 12-18, Београд, 2009.
- М. Петровић, М. Миновић, С. Лазаревић, „**Мобилни уређаји и .NET Web сервиси**“, Зборник радова YU INFO, Копаоник, 2006.
- М. Миновић, В. Штавланин, Д. Старчевић, М. Петровић, „**Примена J2ME технологија у мобилном банкарству**“, часопис ИнфоМ 3(10), 33-37, Београд, 2004.
- С. Лазаревић, М. Петровић, „**Корелација једнакости, хешинга и наслеђивања**“, часопис ИнфоМ 2(8), 37-40, Београд 2003.

Рад објављен на домаћој конференцији

- М. Мухић, М. Петровић, Ј. Томашевић, „**Web сервиси као подршка примени Cost-Benefit анализе**“, Зборник радова YU INFO, Копаоник, 2006.
- С. Лазаревић, М. Петровић, „**Корелација једнакости, хешинга и наслеђивања**“, Зборник радова SYMOPIS, 247-251, Херцег Нови, 2003.

Преглед пројеката

Преглед идејних и главних пројеката и техничких решења (прототип, софтвер) у чијој је изради кандидат учествовао као члан или вођа тима:

- „**PopisPoljoprivrede**“, Републички завод за статистику, Србија, Београд, 2012.

- „*ResursniPortal*“, Министарство здравља, Црна Гора, Подгорица, 2012.
- „*Popis*“, Републички завод за статистику, Србија, Београд, 2011-2012.
- „*Tehnis*“, Министарство економије и регионалног развоја, Србија, Београд, 2010.
- „*BelVille*“, Blok 67 Associates d.o.o., Београд, 2008-2010.
- „*DirectDebit*“, Удружење банака Србије, Београд, 2008-2009.
- „*TiCat*“, Телеком Србија, Београд, 2007-2009.
- „*TB LAB 2006*“, Министарство Здравља, Србија, Београд, 2006-2007.
- „*Cybersure*“, Mantacore Spearhead д.о.о., Београд, 2006-2007.
- „*TB 2005*“, Министарство Здравља, Србија, Београд, 2005.

2. Предмет и циљ истраживања

Проблем који ће бити анализиран у оквиру овог истраживања је развој система складишта података. Узимајући у обзир природу проблема који се решава овим системима може се рећи да проблем истраживања обухвата области као што су: базе података, складишта података, пословна интелигенција и подршка одлучивању.

Савремено пословање је готово немогуће замислити без одговарајуће подршке информационих технологија. Напредак технологије, оштра конкуренција као и честе промене у пословном окружењу утицали су да данашња пословања буду у великој мери зависна од одговарајуће подршке информационих система. Ова подршка се пре свега односи на ефикасно обављање рутинских дневних задатака, односно прикупљање, складиштење и обраду свих података који су неопходни за несметано извршавање дефинисаних пословних процеса.

Ефикасност је потребан, али не и довољан услов успешног пословања. Поред ефикасности, неопходно је постићи и ефективност извршавања пословних процеса. Да би се омогућило постизање ефективности пословања, потребно је обезбедити одговарајућу подршку вишем менаџменту, односно обезбедити ефикасност у процесу доношења пословних одлука. Перманентна анализа пословања и евалуација постављених циљева, као и изналажење нових и бољих начина за остваривање конкурентске предности су услов успешног пословања.

Повећање сложености пословања и све израженија конкуренција која влада у његовом окружењу, повећава притисак на доносиоце одлука (Ponniah, 2011), било да се доносе оперативне, тактичке или стратешке одлуке. Потребне су правовремене и исправне одлуке како би се пословање што брже прилагодило променама и како би се пронашли нови начини стицања конкурентске предности. Оно што карактерише процес доношења одлука у

савременим пословањима је постојање велике количине података и кратки временски рокови у којима је одлуке потребно донети (Suknović & Delibašić, 2010). Имати праве информације, у право време, на правом месту уз разумне трошкове је предуслов доношења исправних пословних одлука (Lenzerini, Vassiliadis, Jarke, & Vassiliou, 2003).

Сложеност проблема који се решавају тактичким и стратешким одлукама, као и значај који те одлуке имају на пословање са једне стране, и немогућност постојећих технологија и система да трансформишу пословне податке у стратешке информације у облику који је погодан за анализе са друге стране, довела је до појаве посебне врсте система који су данас познати под називом пословна интелигенција (енг. *business intelligence*). Појмом пословна интелигенција означавају се савремени системи за подршку одлучивању као посебне гране у информационим технологијама којима је основни циљ унапређење процеса доношења одлука (Suknović & Delibašić, 2010).

У оквиру пословне интелигенције развили су се системи складишта података (енг. *data warehousing systems*), као посебна врста информационих система, који представљају спону између трансакционих и аналитичких система. Потребна за овим системима проистекла је из чињенице да трансакциона и аналитичка обрада података не могу ефикасно да коегзистирају у истом окружењу (Lenzerini, Vassiliadis, Jarke, & Vassiliou, 2003), као и чињенице да постојећи трансакциони системи нису у могућности да обезбеде стратешке информације (Ponniah, 2011).

Системи складишта података треба да омогуће прикупљање пословних података, затим њихову трансформацију у одговарајуће стратешке информације и смештање тако добијених стратешких информација у облику који је погодан за спровођење анализа (Poole, 2003). За прикупљање и трансформацију пословних података спроводи се велики број различитих типова активности које се организују у одговарајуће процесе (енг. *extract-transform-load process*, у даљем тексту ЕТЛ процес), док се за складиштење стратешких информација користе специјализоване базе података које се називају складишта података (енг. *data warehouse*).

У циљу добијања високо квалитетних стратешких информација у облику који је погодан за доношење одлука, потребно је спровести већи број различитих типова функција на подацима у одговарајућем редоследу. Типови функција и редослед у коме се оне извршавају дефинисан је процесом који је у литератури познат под називом ЕТЛ процес. ЕТЛ процес је опште прихваћен назив за процесе у окружењу система складишта података, којима се врши екстраховање и корекција података, затим њихова трансформација и консолидација и на крају учитавање у складиште података (Lazarević, Marjanović, Aničić, & Babarogić, 2010).

Предмет овог истраживања је развој ЕТЛ процеса као најсложенијег и најзначајнијег дела у читавом процесу развоја система складишта података. Основни циљ који се жели постићи овим процесом је правовремено обезбеђивање високо квалитетних стратешких информација у облику који је погодан за доношење пословних одлука. Узимајући ово у обзир, начин на који се пројектују и имплементирају ови процеси у великој мери утиче на квалитет добијених стратешких информација, а самим тим и на употребљивост, односно успех целокупног система складишта података.

Развој ЕТЛ процеса је веома сложен и временски захтеван процес који захтева значајна финансијска средства. У литератури се наводи да је чак 80% времена и уложеног напора у развоју система складишта података управо посвећен развоју ЕТЛ процеса (Golfarelli, Data warehouse life-cycle and design, 2009; Kimball, Ross, Thornthwaite, Mundy, & Becker, 2010). Потребно је велико знање, вештине и способности свих учесника како би се решили бројни проблеми и ризици који владају на пројекту. Поред тога, на пројектима се спроводи велики број различитих типова активности, које заједно са њиховим резултатима треба организовати на одговарајући начин. Све ово указује на неопходност коришћења одговарајућег методолошког приступа, односно формализованог начина на који се ови системи развијају.

У основи методолошких приступа је да омогуће организацијама које се баве развојем да изнова и квалитетно развијају софтверска решења. Они обезбеђују дисциплинован приступ организовању и извођењу пројеката са циљем да се софтверска решења реализују са дефинисаним планом и буџетом, као и да задовољавају дефинисани ниво квалитета. Уопште, методолошким приступима се омогућава савладавање сложености и постизање успеха на пројектима кроз решавање проблема који се првенствено огледају у слабој продуктивности и високим трошковима развоја, али и ниским нивоом квалитета и поузданости решења.

Најсавременији приступ који се предлаже у софтверској индустрији за решавање наведених проблема је развој вођен моделима (енг. *model driven development*, у даљем тексту МДД). Основни циљ који се жели постићи је да се повећа продуктивност и скрати време потребно за развој софтверског решења. За постизање овог циља, МДД прописује издизање нивоа апстракције и коришћење трансформација којима се процес развоја аутоматизује.

У основи овог приступа је коришћење добро познате и доказане технике апстракције, која представља најмоћнији интелектуални алат у савладавању сложености. Наиме, једини методолошки начин савладавања сложености је употреба апстракција, тј. свесно занемаривање детаља на неком нивоу апстракције како би се број концепата са којима се суочавамо редуковао на ниво савладиве сложености.

У складу са тим, овим приступом се прописује коришћење модела као примарних софтверских артефаката којима се води развој софтвера. Развој започиње израдом модела на вишем нивоу апстракције (енг. *platform independent model*, у даљем тексту ПИМ), а затим се кроз низ сукцесивних профињавања прелази на ниже нивое апстракције (енг. *platform specific model*, у даљем тексту ПСМ), све док се не стигне до изворног кода. Ово профињавање, односно постепено увођење детаља, се имплементира помоћу одговарајућих трансформација које се такође специфицирају преко модела и чије је извршавање аутоматизовано.

Повећање нивоа апстракције води ка повећању продуктивности, али и ка постизању вишег нивоа квалитета решења. Поред тога, примена трансформација омогућава аутоматизацију процеса развоја, чиме се додатно утиче на повећање продуктивности и постизање ефикасности у развоју софтверских решења.

На основу свега наведеног, циљ овог истраживања је да се у складу са доступном литературом сагледа проблем развоја ЕТЛ процеса, затим идентификују и анализирају релевантни приступи, као и да се истражи могућност унапређења, односно аутоматизације процеса развоја ЕТЛ процеса применом МДД принципа.

Оквирна листа релевантних библиографских извора која ће се користити приликом израде докторске дисертације:

- Adamson, C. (2010). *Star Schema The Complete Reference*. McGraw-Hill Education.
- Akkaoui, Z., Mazón, J.-N., Vaisman, A., & Zimányi, E. (2012). BPMN-Based Conceptual Modeling of ETL Processes. 7448, str. 1-14.
- Akkaoui, Z., Zimányi, E., Mazón, J.-N., & Trujillo, J. (2011). A model-driven framework for ETL process development., (str. 45--52).
- Albin, S. T. (2003). *The art of software architecture: design methods and techniques*. Wiley Pub.
- Ambler, S. W. (2006, Novembar). *A RUP-based approach to developing a data warehouse - Part 1: Setting the stage*. Preuzeto Septembar 2012 sa IBM developerWorks: <http://www.ibm.com/developerworks/rational/library/nov06/ambler/index.html>
- Ambler, S. W. (2006, Novembar). *A RUP-based approach to developing a data warehouse - Part 2: Building the data warehouse, one iteration at a time*. Preuzeto Septembar 2012 sa IBM developerWorks: <http://www.ibm.com/developerworks/rational/library/dec06/ambler/>
- Aničić, N., Nešković, S., Vučković, M., & Cvetković, R. (2012). Specification of data schema mappings using weaving models. *Computer Science and Information Systems*, 9, 539-559.
- Ariyachandra, T., & Watson, H. (2010). Key organizational factors in data warehouse architecture selection. *Decision Support Systems*, 49, 200-212.
- Atkinson, C., & Kuhne, T. (2003, sep). Model-Driven Development: A Metamodeling Foundation. *Software, IEEE*, 20, 36--41.
- Babarogić, S. (2004). *Primena arhitekture bazirane na modelima na razvoj programskih sistema u .NET okruženju*. Beograd: Fakultet organizacionih nauka, Univerzitet u Beogradu.
- Ballard, C. (2006). *Dimensional Modeling: In a Business Intelligence Environment*. Vervante.
- Bezivin, J. (2004). In Search of a Basic Principle for Model Driven Engineering. *UPGRADE -- The European Journal for the Informatics Professional*, 5, 21--24.
- Böhnlein, M., & Ulbrich-vom Ende, A. (2000). Business process oriented development of data warehouse structures. *Jung, R., Winter, R.(Eds.): Data Warehousing*, 3-21.
- Bezivin, J. (2005). On the unification power of models. *Software & Systems Modeling*, 4(2), 171-188.

- Casters, M., Bouman, R., & van Dongen, J. (2010). *Pentaho Kettle Solutions: Building Open Source ETL Solutions with Pentaho Data Integration*. Wiley.
- Clark, T., Sammut, P., & Willans, J. (2008). Applied metamodeling: a foundation for language driven development.
- Corr, L., & Stagnitto, J. (2011). *Agile Data Warehouse Design: Collaborative Dimensional Modeling, from Whiteboard to Star Schema*. DecisionOne Press.
- Didonet, M., Fabro, D., Bézivin, J., & Valduriez, P. (2006). Weaving Models with the Eclipse AMW plugin.
- El-Sappagh, S. H., Hendawi, A. M., & Bastawissy, A. H. (2011). A proposed model for data warehouse ETL processes. *Journal of King Saud University - Computer and Information Sciences*, 23(2), 91-104.
- Frankel, D. (2002). *Model Driven Architecture: Applying MDA to Enterprise Computing*. John Wiley & Sons, Inc.
- Giordano, A. D. (2011). *Data Integration Blueprint and Modeling: Techniques for a Scalable and Sustainable Architecture*. IBM Press,.
- Golfarelli, M. (2009). Data warehouse life-cycle and design. *Collection of Encyclopedia of Database Systems*, 658--664.
- Golfarelli, M. (2010). From User Requirements to Conceptual Design in Warehouse Design: A Survey. U L. Bellatreche, *Data Warehousing Design and Advanced Engineering Applications: Methods for Complex Construction* (str. 1-16). Information Science Reference.
- Golfarelli, M., & Rizzi, S. (2009). *Data Warehouse Design: Modern Principles and Methodologies*. McGraw-Hill, Inc.
- Gong, L., Olivas, M., Posluszny, C., Posluszny, C., & McMillan, G. (2005, 6 16). *Deliver an effective and flexible data warehouse solution, Part 1: Engage your customers in planning the data warehouse project*. Preuzeto 09 15, 2012 sa IBM developerWorks: <http://www.ibm.com/developerworks/data/library/techarticle/dm-0506gong/>
- Gong, L., Olivas, M., Posluszny, C., Posluszny, C., & McMillan, G. (2005, 7 14). *Deliver an effective and flexible data warehouse solution, Part 2: Develop a warehouse data model*. Preuzeto 9 15, 2012 sa IBM developerWorks: <http://www.ibm.com/developerworks/data/library/techarticle/dm-0507gong/>
- Gong, L., Olivas, M., Posluszny, C., Posluszny, C., & McMillan, G. (2005, 7 4). *Deliver an effective and flexible data warehouse solution, Part 3: Design and implement a warehouse ETL process*. Preuzeto 9 15, 2012 sa IBM developerWorks: <http://www.ibm.com/developerworks/data/library/techarticle/dm-0508gong/>
- Greenfield, J., & Short, K. (2003). Software factories: assembling applications with patterns, models, frameworks and tools. (str. 16--27). ACM.
- Griesemer, B. (2009). *Oracle Warehouse Builder 11g: Getting Started*. Packt Publishing, Limited.
- Inmon, W. H. (2005). *Building the Data Warehouse*. Wiley.
- Jiang, B. (2011). *Constructing Data Warehouses with Metadata-Driven Generic Operators, and More*. Dbj Publishing.
- Jovanovic, V., & Bojicic, I. (2012). Conceptual Data Vault Model. *Proceedings of the SAIS Conference*, 23, str. 131-136.
- Jovanovic, V., Bojicic, I., Knowles, C., & Pavlic, M. (2012). Persistent staging area models for data warehouses. *Issues in Information Systems*, 13, 121-132.
- Kimball, R., & Ross, M. (2002). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. Wiley.
- Kimball, R., Ross, M., Thornthwaite, W., Mundy, J., & Becker, B. (2010). *The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence*. Wiley.
- Kleppe, A. G., Warmer, J., & Bast, W. (2003). *MDA Explained: The Model Driven Architecture: Practice and Promise*. Addison-Wesley Longman Publishing Co., Inc.
- Kuhne, T. (2006). Matters of (Meta-)Modeling. *Software and Systems Modeling*, 5, 369-385.
- Kurtev, I., Bezivin, J., Jouault, F., & Valduriez, P. (2006). Model-based DSL frameworks. (str. 602--616). ACM.
- Lazarević, B., Marjanović, Z., Aničić, N., & Babarogić, S. (2010). *Baze podataka*. Beograd: Fakultet organizacionih nauka, Univerzitet u Beogradu.

- Lenzerini, M., Vassiliadis, P., Jarke, M., & Vassiliou, Y. (2003). *Fundamentals of Data Warehouses*. Springer.
- Linstedt, D. E. (2002, Jul). *Data Vault Series 1 - Data Vault Overview*. Preuzeto Jun 2012 sa The Data Administration Newsletter: <http://www.tdan.com/view-articles/5054/>
- Linstedt, D. E. (2003, Januar). *Data Vault Series 2 - Data Vault Components*. Preuzeto Jun 2012 sa The Data Administration Newsletter: <http://www.tdan.com/view-articles/5155/>
- Linstedt, D. E. (2003, April). *Data Vault Series 3 - End Dates and Basic Joins*. Preuzeto Jun 2012 sa The Data Administration Newsletter: <http://www.tdan.com/view-articles/5067/>
- Linstedt, D. E. (2004, Januar). *Data Vault Series 4 - Link Tables*. Preuzeto Jun 2012 sa The Data Administration Newsletter: <http://www.tdan.com/view-articles/5172/>
- Linstedt, D. E. (2005, Januar). *Data Vault Series 5 - Loading Practices*. Preuzeto Jun 2012 sa The Data Administration Newsletter: <http://www.tdan.com/view-articles/5285/>
- List, B., Bruckner, R., Machaczek, K., & Schiefer, J. (2002). A comparison of data warehouse development methodologies case study of the process warehouse., (str. 203--215).
- Luján-Mora, S., & Trujillo, J. (2004). A Data Warehouse Engineering Process. 3261, str. 14-23. Springer.
- Luján-Mora, S., Vassiliadis, P., & Trujillo, J. (2004). Data Mapping Diagrams for Data Warehouse Design with UML. 3288, str. 191-204.
- Mazón, J.-N., & Trujillo, J. (2008). An MDA approach for the development of data warehouses. *Decision Support Systems*, 45, 41--58.
- Milicev, D. (2009). *Model-Driven Development with Executable UML*. Wiley.
- Mukerji, J., & Miller, J. (2003, Јун). *MDA Guide Version 1.0.1*. Преузето Октобар 2012 са OMG: <http://www.omg.org/cgi-bin/doc?omg/03-06-01>
- Mundy, J., Thornthwaite, W., & Kimball, R. (2011). *The Microsoft Data Warehouse Toolkit: With SQL Server 2008 R2 and the Microsoft Business Intelligence Toolset*. Wiley.
- Muñoz, L., Mazón, J.-N., & Trujillo, J. (2009). Automatic generation of ETL processes from conceptual models., (str. 33--40).
- Muñoz, L., Mazón, J.-N., Pardillo, J., & Trujillo, J. (2008). Modelling ETL Processes of Data Warehouses with UML Activity Diagrams. 5333, str. 44-53.
- Object Management Group. (2003). *Common Warehouse Metamodel 1.1 Specification*. Object Management Group, Inc.
- Pahwa, P., Taneja, S., & Thakur, G. (2011). Uclean: A Requirement Based Object-Oriented ETL Framework. *International Journal of Computer Science & Engineering Survey (IJCSSES)*, 2(4), 45-55.
- Petrović, M. (2012). *Uporedna analiza izabranih metodologija razvoja softvera*. Seminarski rad, Fakultet organizacionih nauka, Univerzitet u Beogradu, Beograd.
- Ponniah, P. (2011). *Data Warehousing Fundamentals for IT Professionals*. Wiley.
- Poole, J. (2003). *Common Warehouse Metamodel Developers Guide*. John Wiley & Sons.
- Reeves, L. (2009). *A Managers Guide to Data Warehousing*. Wiley.
- Rizzi, S. (2008). Business Intelligence. U B. W. Wah (Ur.), *Wiley Encyclopedia of Computer Science and Engineering*. John Wiley & Sons, Inc.
- Rizzi, S. (2008). Data Warehouse. U B. W. Wah (Ur.), *Wiley Encyclopedia of Computer Science and Engineering*. John Wiley & Sons, Inc.
- Romero, O., & Abelló, A. (2009). A survey of multidimensional modeling methodologies. *International Journal of Data Warehousing and Mining (IJDWM)*, 5, 1-23.
- Romero, O., & Abello, A. (2010). A framework for multidimensional design of data warehouses from ontologies. *Data & Knowledge Engineering*, 69, 1138-1157.
- Simitsis, A. (2005). Mapping conceptual to logical models for ETL processes., (str. 67--76).
- Simitsis, A., & Vassiliadis, P. (2003). A methodology for the conceptual modeling of ETL processes.
- Simitsis, A., & Vassiliadis, P. (2008). A method for the mapping of conceptual designs to logical blueprints for ETL processes. *Decision Support Systems*, 45, 22--40.
- Simitsis, A., Vassiliadis, P., Terrovitis, M., & Skiadopoulos, S. (2005). Graph-based modeling of ETL activities with multi-level transformations and updates. str. 43--52.
- Suknović, M., & Delibašić, B. (2010). *Poslovna inteligencija i sistemi za podršku odlučivanju*. Beograd: Fakultet organizacionih nauka, Univerzitet u Beogradu.

- Team, M. P. (2009). *Microsoft Registered Application Architecture Guide*. Microsoft Press.
- Trujillo, J., & Luján-Mora, S. (2003). A UML Based Approach for Modeling ETL Processes in Data Warehouses. *2813*, str. 307-320.
- Vassiliadis, P. (2009). A Survey of Extract-Transform-Load Technology. *International Journal of Data Warehousing and Mining (IJDWM)*, 5, 1-27.
- Vassiliadis, P., Simitsis, A., & Baikousi, E. (2009). A taxonomy of ETL activities. (str. 25-32). ACM.
- Vassiliadis, P., Simitsis, A., & Skiadopoulos, S. (2002). Conceptual modeling for ETL processes., (str. 14-21).
- Vassiliadis, P., Simitsis, A., & Skiadopoulos, S. (2002). Modeling ETL activities as graphs., (str. 52-61).
- Vassiliadis, P., Simitsis, A., Georgantas, P., & Terrovitis, M. (2003). A Framework for the Design of ETL Scenarios., (str. 520-535).
- Vassiliadis, P., Simitsis, A., Georgantas, P., Terrovitis, M., & Skiadopoulos, S. (2005). A generic and customizable framework for the design of ETL scenarios. *Information Systems*, 30, 492-525.
- Vassiliadis, P., Vagena, Z., Skiadopoulos, S., Karayannidis, N., & Sellis, T. (2001). ARKTOS: towards the modeling, design, control and execution of ETL processes. *Information Systems*, 26, 537-561.
- Vuckovic, M., Petrovic, M., Turajlic, N., & Stanojevic, M. (2012, DEC). The Specification of ETL Transformation Operations based on Weaving Models. *INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL*, 7, 968-975.
- Watson, H. J., & Ariyachandra, T. (2005). Data Warehouse Architectures: Factors in the Selection Decision and the Success of the Architectures. *Terry College of Business, University of Georgia*.
- Winter, R., & Strauch, B. (2003). A method for demand-driven information requirements analysis in data warehousing projects., (str. 9--pp).

3. Полазне хипотезе

На основу анализе доступне литературе и постављеног предмета и циља истраживања може се поставити:

Општа хипотеза

- Могућа је аутоматизација развоја ЕТЛ процеса коришћењем МДД приступа.

Посебне хипотезе

- Могуће је дефинисати специфичан језик (метамодел) за спецификацију тока ЕТЛ процеса.
- Могуће је дефинисати специфичан језик (метамодел) за спецификацију скупа операција трансформација ЕТЛ процеса.
- Могуће је дефинисати аутоматске трансформације модела, заснованих на дефинисаним специфичним језицима, на основу којих се може аутоматски генерисати изворни код за изабрану ЕТЛ платформу.
- Предложеним решењем могуће је унапредити постојећу индустријску праксу.

4. Научне методе истраживања

За успешну реализацију истраживања у раду ће бити коришћене опште методе истраживања, као што су: аналитичко-синтетички метод, индуктивно-дедуктивни метод, метод апстракције и конкретизације, метод генерализације и специјализације и метод моделовања.

За спецификацију модела и спецификацију њихових трансформација користиће се неколико приступа моделовању: директно моделовање, интеграција подмодела, метамоделовање и конкретизација генеричких модела.

5. Очекивани научни допринос

Резултати који ће проистећи из истраживања у овој докторској дисертацији ће садржати већи број доприноса, међу којима се издвајају:

Научни доприноси

- Преглед и компарација резултата досадашњих истраживања из области развоја складишта података.
- Анализа и систематизација операција трансформација података у ЕТЛ процесима.
- Дефинисање специфичног језика (метамодел) за спецификацију тока ЕТЛ процеса.
- Дефинисање специфичног језика (метамодел) за спецификацију операција трансформација ЕТЛ процеса.
- Дефинисање специфичног архитектуралног патерна, односно логичке софтверске архитектуре ЕТЛ процеса.
- Дефинисање модела трансформација, заснованих на дефинисаном језику за операције трансформације ЕТЛ процеса, на основу којих се могу добити извршне трансформације за изабрану ЕТЛ платформу.

Стручни доприноси

- Реализација трансформација модела, базираних на дефинисаним моделима трансформација, који аутоматски генеришу извршни код за изабрану ЕТЛ платформу.
- Уводи се и биће развијен фрејмворк који треба технолошки да подржи дефинисане језике специфичне намене и омогући што већу аутоматизацију процеса развоја.

6. План истраживања и структура рада

У уводном поглављу ће бити детаљно приказан предмет и циљ истраживања. Затим ће бити приказани идентификовани проблеми дате области и биће дат кратак опис структуре дисертације по поглављима.

Другим поглављем ће се дати теоријски оквир истраживања са циљем да се опишу и ближе одреде појмови из области којом се бави истраживање. Пре свега, разматраће се појмови пословне интелигенције и система складишта података, али и принципи најсавременијег приступа развоју софтвера вођеног моделима.

Треће поглавље ће се бавити проблемом ЕТЛ процеса. У првом делу ће се, на основу доступне литературе, описати основни елементи ЕТЛ процеса, а затим ће се дати преглед и анализа постојећих приступа у области. У другом делу поглавља ће се представити свеобухватна визија, односно предложено решење. Елементи предложеног решења ће у поглављима која следе бити детаљно разматрани.

У четвртом поглављу ће се разматрати проблем спецификације ЕТЛ процеса. Детаљно ће се описати предложени концептуални модел ЕТЛ процеса. Извршиће се анализа и систематизација операција трансформација (таксономија), као предуслова за дефинисање специфичног језика (метамодела) за спецификацију операција трансформација. Поред тога, дефинисаће се и специфичан језик (метамодел) за спецификацију тока извршавања ЕТЛ процеса.

Петим поглављем ће се разматрати проблем реализације спецификације ЕТЛ процеса. Разматраће се савремени архитектурални патерни и у складу са тим дефинисаће се специфична логичка софтверска архитектура погодна за софтверска решења из домена ЕТЛ процеса. Идентификоваће се неопходни типови софтверских компоненти, њихове улоге и одговорности, као и интерфејси неопходни за комуникацију међу њима.

Шестим поглављем ће се бавити проблемом имплементације ЕТЛ процеса. Разматраће се развој (имплементација) специфичног фрејмворка, односно одговарајуће извршне ЕТЛ платформе као основе за имплементацију ЕТЛ процеса. Поред тога, разматраће се и проблем оптимизације извршавања ЕТЛ процеса, тј. дефинисања оптималног плана извршавања процеса у односу на специфициране операције трансформација и физичку архитектуру софтверског решења.

У седмом поглављу ће бити дат опис прототипа фрејмворка развијеног у *Microsoft .NET* технолошком окружењу, док ће у осмом поглављу бити представљена експериментална валидација предложеног решења. На репрезентативном примеру из праксе ће бити валидирано предложено решење и указане предности у односу на постојећу праксу.

Закључак, научни доприноси и правци даљих истраживања биће разматрани у деветом поглављу.

На крају рада ће бити дат списак коришћене литературе и одговарајући прилози.



7. Закључак и предлог

На основу образложења предложене докторске дисертације, Комисија закључује да кандидат, **Марко Петровић**, дипломирани инжењер, испуњава све услове предвиђене Законом о високом образовању и Статутом Факултета, за израду докторске дисертације, с обзиром на елементе изнете у овом извештају. Према мишљењу Комисије, тема „**РАЗВОЈ ПРОЦЕСА ЕКСТРАКЦИЈЕ, ТРАНСФОРМАЦИЈЕ И ПУЊЕЊА ПОДАТАКА СКЛАДИШТА ПОДАТАКА ЗАСНОВАН НА МОДЕЛОМ ВОЂЕНОМ ПРИСТУПУ**“ спада у научно подручје које се развија на Факултету организационих наука. Тема докторске дисертације по предмету истраживања, садржају и очекиваним доприносима представља значајно подручје истраживања, како са теоријског, тако и са аспекта примене у пракси, у области Техничких наука, уже научне области Информациони системи.

На основу наведеног, Комисија предлаже Наставно-научном већу Факултета организационих наука да кандидату **Марку Петровићу** одобри израду докторске дисертације под насловом „**РАЗВОЈ ПРОЦЕСА ЕКСТРАКЦИЈЕ, ТРАНСФОРМАЦИЈЕ И ПУЊЕЊА ПОДАТАКА СКЛАДИШТА ПОДАТАКА ЗАСНОВАН НА МОДЕЛОМ ВОЂЕНОМ ПРИСТУПУ**“ и за ментора се именује проф. др Зоран Марјановић, редовни професор Факултета организационих наука.

ЧЛАНОВИ КОМИСИЈЕ

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

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

Проф. др Милица Вучковић, ванредовни професор,
Факултет организационих наука, Универзитет у Београду

др Слађан Бабарогић, доцент,
Факултет организационих наука, Универзитет у Београду

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