

УНИВЕРЗИТЕТ У БЕОГРАДУ
ФАКУЛТЕТ ОРГАНИЗАЦИОНИХ НАУКА

АЛЕКСАНДАР М. СИМОВИЋ

МОДЕЛ ПАМЕТНЕ БИБЛИОТЕКЕ ЗАСНОВАН НА *BIG*

***DATA* ТЕХНОЛОГИЈАМА**

ДОКТОРСКА ДИСЕРТАЦИЈА

БЕОГРАД, 2020

UNIVERSITY OF BELGRADE
FACULTY OF ORGANIZATIONAL SCIENCES

ALEKSANDAR M. SIMOVIĆ

SMART LIBRARY MODEL BASED ON BIG DATA

TECHNOLOGIES

DOCTORAL DISSERTATION

BELGRADE, 2020

Ментор:

проф. др Александра Лабус, ванредни професор
Факултет организационих наука Универзитета у Београду

Чланови комисије:

проф. др Зорица Богдановић, ванредни професор
Факултет организационих наука Универзитета у Београду

проф. др Саша Лазаревић, ванредни професор
Факултет организационих наука Универзитета у Београду

проф. др Вељко Јеремић, ванредни професор
Факултет организационих наука Универзитета у Београду

проф. др Наташа Бојковић, ванредни професор
Саобраћајни факултет Универзитета у Београду

Датум одбране: _____ 2020. године.

Модел паметне библиотеке заснован на big data технологијама

Апстракт:

Предмет истраживања докторске дисертације је развој модела паметне библиотеке заснованог на big data технологијама и сервисима. Централни истраживачки проблем разматран у раду је развој big data инфраструктуре и сервиса паметне библиотеке који омогућавају интелигентну претрагу и препоруку библиотечког садржаја. Посебан циљ рада је да испита могућност интеграције развијеног модела са паметним образовним окружењима у циљу унапређења квалитета образовног процеса.

У докторској дисертацији је представљен модел паметне библиотеке као интегралног дела образовног система који може да побољша квалитет и свеобухватност наставних ресурса и повећа мотивацију у процесу учења препоручивањем садржаја од интереса. Модел описан у раду омогућава примену big data система за анализу, обраду и визуализацију података прикупљених из различитих извора и обухвата њихову интеграцију у паметну библиотеку. Циљ развоја паметних библиотека је да се унапреде библиотечки пословни процеси и да се корисницима пруже иновативни сервиси за претрагу и коришћење садржаја.

У дисертацији се разматрају различите перспективе имплементације big data решења за паметне библиотеке као део континуираног образовног процеса, са посебним фокусом на интеграцију традиционалних система и big data технологија. Поред наведених компонента система, модел обухвата инфраструктуру и интеграцију система препоруке колаборативног филтрирања извора различитих података са big data технологијама.

Модел је евалуиран кроз тестирање и мерење релевантних параметара перформанси који утичу на ефикасност предложеног модела.

Кључне речи: big data, системи препорука, анализа података, складиштење података, паметне библиотеке, електронско пословање, електронско образовање

Научна област: Информациони системи и технологије

Ужа научна област: Електронско пословање

Smart library model based on big data technologies

Abstract:

The subject of this doctoral dissertation research is the development of a smart library model based on big data technologies and services. The central research problem discussed in the thesis is the development of big data infrastructure and smart library services that enable intelligent searches and recommendations from the library content. A particular focus of the paper is an examination of the possibility of integrating the developed model into a smart educational environment in order to improve the quality of the educational process.

The thesis presents a model of the smart library as an integral part of the educational system that would improve quality level and comprehensiveness of learning resources and increase the motivation of its users through content-aware recommendations. The model described in the thesis considers the possibilities of applying a big data system for the collection, analysis, processing and visualization of data from multiple sources, and the integration of data into the smart library. The goal of developing a smart library is to improve the library's business process and to offer users innovative methods to search and content use.

The thesis discusses the perspective of the implementation of a big data solution for smart libraries as a part of a continuous learning process, with the aim of improving the results of library operations by integrating traditional systems with big data technology. In addition to the above system components, the model includes the infrastructure and integration of a recommender system for collaborative filtering by incorporating multiple sources of differential data with big data technologies.

Within the evaluation of the model, testing and measurement of the relevant performance parameters which influence the efficiency of the proposed model were carried out.

Keywords: *big data, recommender systems, data analysis, data storage, smart libraries, e-business, e-education*

Scientific field: *Information systems and technologies*

Narrow scientific field: *E-business*

САДРЖАЈ:

1.	УВОД.....	10
1.1.	ДЕФИНИСАЊЕ ПРЕДМЕТА ИСТРАЖИВАЊА.....	10
1.2.	ЦИЉЕВИ ИСТРАЖИВАЊА	12
1.3.	ПОЛАЗНЕ ХИПОТЕЗЕ	12
1.4.	МЕТОДЕ ИСТРАЖИВАЊА.....	13
1.5.	СТРУКТУРА И ОРГАНИЗАЦИЈА РАДА	13
2.	АНАЛИЗА ПОСТОЈЕЋИХ РЕШЕЊА И ОСВРТ НА ЛИТЕРАТУРУ	15
2.1.	ТРАДИЦИОНАЛНЕ И ДИГИТАЛНЕ БИБЛИОТЕКЕ	16
2.2.	СИСТЕМИ ПРЕПОРУКА У BIG DATA ОКРУЖЕЊУ	17
2.3.	ПАМЕТНЕ БИБЛИОТЕКЕ	20
3.	ТЕХНОЛОГИЈЕ ПРИМЕЊЕНЕ ЗА РАЗВОЈ МОДЕЛА	28
3.1.	BIG DATA	28
3.1.1.	<i>BIG DATA</i> КОНЦЕПТИ	29
3.1.2.	<i>BIG DATA</i> ИНФРАСТРУКТУРА	31
3.1.3.	<i>BIG DATA</i> ИНФРАСТРУКТУРА <i>MICROSOFT</i>	35
3.1.4.	<i>BIG DATA</i> СЕРВИСИ	39
3.1.5.	<i>BIG DATA</i> АНАЛИТИКА.....	44
3.2.	СИСТЕМИ ПРЕПОРУКА.....	49
3.2.1.	СИСТЕМИ БАЗИРАНИ НА САДРЖАЈУ	50
3.2.2.	СИСТЕМИ КОЛАБОРАТИВНОГ ФИЛТРИРАЊА.....	52
3.2.3.	ХИБРИДНИ СИСТЕМИ	58
3.2.4.	ПЕРСОНАЛИЗАЦИЈА СИСТЕМА ПРЕПОРУКА.....	59
3.2.5.	ИНТЕЛИГЕНТНИ АГЕНТИ У СИСТЕМИМА ПРЕПОРУКА	61
3.2.6.	РЕАЛИЗАЦИЈА СИСТЕМА ПРЕПОРУКЕ	62
3.2.7.	КОМПАРАТИВНА АНАЛИЗА СИСТЕМА ПРЕПОРУКА	66
4.	МОДЕЛ ПАМЕТНЕ БИБЛИОТЕКЕ ЗАСНОВАН НА <i>BIG DATA</i> ТЕХНОЛОГИЈАМА	74
4.1.	ПРОЈЕКТОВАЊЕ АРХИТЕКТУРЕ МОДЕЛА ПАМЕТНЕ БИБЛИОТЕКЕ ЗАСНОВАНОГ НА <i>BIG DATA</i> ТЕХНОЛОГИЈАМА	75
4.1.1.	СТРУКТУРА МОДЕЛА ПАМЕТНЕ БИБЛИОТЕКЕ	77
4.1.2.	ПРОЈЕКТОВАЊЕ ХАРДВЕРСКОГ СИСТЕМА ПАМЕТНЕ БИБЛИОТЕКЕ	78
4.1.3.	СОФТВЕРСКИ СЕРВИСИ ПРОЈЕКТОВАНОГ МОДЕЛА ПАМЕТНЕ БИБЛИОТЕКЕ.....	84
4.1.4.	АПЛИКАЦИЈЕ СИСТЕМА ПАМЕТНЕ БИБЛИОТЕКЕ	94
4.2.	ИНТЕГРАЦИЈА ПОДАКА У ЕКОСИСТЕМ.....	100

4.2.1.	ПОДАЦИ ИЗ ИНФОРМАЦИОНОГ СИСТЕМА ОБРАЗОВНЕ ИНСТИТУЦИЈЕ	100
4.2.2.	БИБЛИОТЕЧКИ СКУП ПОДАТАКА	100
4.2.3.	ПОДАЦИ СА <i>LMS MOODLE</i> ПЛАТФОРМЕ	100
4.2.4.	ЛИСТА ТАБЕЛА УЧИТАНИХ СКУПОВА ПОДАТАКА	100
4.2.5.	ИНТЕГРАЦИЈА ПОДАТАКА ДРУШТВЕНИХ МЕДИЈА	101
4.2.6.	ПРИМЕНА <i>IoT</i> ТЕХНОЛОГИЈА У МОДЕЛУ ПАМЕТНЕ БИБЛИОТЕКЕ	103
4.3.	МЕРЕЊЕ ПЕРФОРМАНСИ МОДЕЛА	105
5.	ПРИМЕНА РАЗВИЈЕНОГ МОДЕЛА	106
5.1.	ПЛАН ПРИМЕНЕ РАЗВИЈЕНОГ МОДЕЛА	106
5.2.	ИМПЛЕМЕНТАЦИЈА РЕШЕЊА	106
5.3.	ЕВАЛУАЦИЈА РЕШЕЊА	110
5.4.	АНАЛИЗА ПОСТИГНУТИХ РЕЗУЛТАТА	114
6.	ДИСКУСИЈА	117
6.1.	ОГРАНИЧЕЊА	120
6.2.	ИСТРАЖИВАЧКЕ ИМПЛИКАЦИЈЕ	120
6.3.	ПРАКТИЧНЕ ИМПЛИКАЦИЈЕ	122
7.	НАУЧНИ И СТРУЧНИ ДОПРИНОСИ	124
8.	БУДУЋА ИСТРАЖИВАЊА	126
9.	ЗАКЉУЧАК	127
10.	ЛИТЕРАТУРА	128
11.	ПРИЛОЗИ	148
11.1.	СПИСАК СЛИКА	157
11.2.	СПИСАК ТАБЕЛА	159
11.3.	БИОГРАФСКИ ПОДАЦИ О КАНДИДАТУ	160
11.4.	ИЗЈАВА О АУТОРСТВУ	163
11.5.	ИЗЈАВА О ИСТОВЕТНОСТИ ШТАМПАНЕ И ЕЛЕКТРОНСКЕ ВЕРЗИЈЕ ДОКТОРСКОГ РАДА	164
11.6.	ИЗЈАВА О КОРИШЋЕЊУ	165

1. УВОД

1.1. ДЕФИНИСАЊЕ ПРЕДМЕТА ИСТРАЖИВАЊА

Предмет истраживања докторске дисертације је развој модела паметне библиотеке заснованог на *big data* технологијама. Централни истраживачки проблем разматран у дисертацији је развој *big data* инфраструктуре и сервиса паметне библиотеке који омогућавају интелигентну претрагу и препоруку библиотечког садржаја. Посебан циљ рада је да испита могућност интеграције развијеног модела са паметним образовним окружењима у циљу унапређења квалитета образовног процеса.

Паметна библиотека је нови модел библиотеке опремљене интелигентним уређајима као што су сензори, камере, уређаји за препознавање лица и говора, RFID тагови и читачи, и други (Radenković et al., 2017). Циљ развоја паметних библиотека је да се унапреде библиотечки пословни процеси и да се корисницима пруже иновативни сервиси за претрагу и коришћење садржаја (Shiwei, 2012; Liu and Shen, 2018). Савремени библиотечки системи обухватају велике количине штампаних, мултимедијалних и дигитализованих материјала из различитих извора (Hey et al., 2009; Bekkerman et al., 2011; Rubin, 2017). Складиштење, аналитичка обрада и визуализација су комплексни и не могу се ефикасно обављати применом традиционалних библиотечких информационих система (Eliot and Rose, 2009; Ahrens et al., 2011). Бројна истраживања указују на проблеме приликом обраде великих количина сложених типова података и предлажу примену *big data* технологија за решавање ових проблема (Cukier and Mayer-Schoenberger, 2013; Horstmann and Brase, 2016; Bhat, 2018).

У савременој литератури, *big data* представља важну технолошку и научно-истраживачку област (Panetta, 2016). Бројне научно-истраживачке институције и компаније у јавном и приватном сектору развијају иновативне моделе за складиштење, претрагу и анализу великих количина података (Fey et al., 2008; Boyd and Crawford, 2012). Примарни задаци *big data* технологија су: складиштење података који се великом брзином генеришу кроз динамичку интеракцију корисника; омогућити претрагу и анализу различитих формата некомпатибилних и неконзистентних структура података; и омогућити подршку за доношење одлука у реалном времену (Laneu, 2001; Vuuya et al., 2016).

Велики број научних радова и научно-истраживачких пројеката бави се развојем *big data* инфраструктуре и сервиса у различитим доменима пословања (Chen and Chao, 2008; Romero et al., 2009; Lu et al., 2011; Santos et al., 2014; Sun et al., 2014; He et al., 2014; Burnap et al., 2014; Raghupathi and Raghupathi, 2014; Thaduri et al.,

2015; Yao et al., 2015; Duan and Xiong, 2015; Ismail and Al-Feel, 2015; Chandio et al., 2015; Saboo et al., 2016; De Meo et al., 2017; Khan et al., 2015; Yi et al., 2017; Khan et al., 2017; Amato et al., 2017; Rodrigues et al., 2018). Међутим, анализа научне литературе указује на недостатак свеобухватног истраживања у области примене *big data* сервиса у паметним библиотекама. Такође, мали број истраживања разматра ово питање из перспективе образовне институције (Simović, 2018).

Библиотечки системи образовних институција су специфични јер треба да представљају саставни део образовног процеса и да буду интегрисани са другим образовним сервисима. Примарни циљ примене *big data* технологија у библиотечком пословању образовних институција је могућност стварања паметног библиотечког екосистема за управљање великим количинама података из различитих извора, као и њихово сакупљање, повезивање, интегрисање, складиштење, анализирање и визуализовање. Осим података о библиотечкој грађи, библиотечки информациони системи образовних институција треба да садрже довољно информација о студентима како би се адекватно одговорило на упит студента и препоручила релевантна литература.

Системи препоруке се у великој мери могу унапредити применом *big data* концепта и заузимају место широко примењених алата у разним областима електронског пословања (Philip Chen and Zhang, 2014). Циљ система препоруке је да смањи преоптерећење информација (енгл. *Information Overload*) тако што извлачи релевантан садржај из велике количине података (Toffler, 1970; Gris , and Gallupe, 1999; Eppler and Mengis, 2004). Према техникама и алгоритмима који се користе при процени корисности садржаја, основни системи препоруке се деле на: системе препоруке базиране на садржају (енгл. *Content Based Recommendations*); системе колаборативног филтрирања (енгл. *Collaborative Filtering*); и хибридне системе препорука (енгл. *Hybrid Recommender Systems*) (Herlocker et al., 2000; Adomavicius and Tuzhilin, 2005). Ови системи предвиђају ниво интересовања корисника или корисности одређеног садржаја и рангирају их према предвиђеним вредностима (Bernardes et al., 2014). Системи базирани на садржају се заснивају на претпоставци да ће корисност садржаја бити слична оној коју је корисник у прошлости бирао, док се код колаборативног филтрирања, корисност предвиђа на основу процене садржаја, путем куповине или оцењивања, од стране других корисника система (Lu et al., 2015). Комбинацијом ових алгоритама ствара се хибридни систем препорука који израчунава и оцену корисника и функцију садржаја (Lin, 2014). Системи препоруке имају велики потенцијал примене у библиотечким информационим системима и могу да допринесу бољем разумевању потреба студената предлагањем садржаја од интереса (Simović, 2018).

1.2. ЦИЉЕВИ ИСТРАЖИВАЊА

Примарни циљ докторске дисертације је развој и евалуација модела паметне библиотеке заснованог на *big data* технологијама. Дисертација обухвата релевантна теоријска и практична решења у домену *big data* технологија и система препоруке у паметним библиотекама. Предложени модел паметне библиотеке интегрише више извора различитих података у *big data* технолошки оквир. У образовном контексту, интеграција података обухватиће различите изворе: податке из система за управљање учењем; евиденције сервера електронске продавнице књига; податке о коришћењу књига из библиотеке; податке прикупљене са сензора у паметном образовном окружењу; и податке са друштвених мрежа. Резултати евалуације система приказаће економске и техничке параметре перформанси развијеног модела, као и образовне параметре перформанси за процену задовољства студената и утицаја примене развијеног система на исходе учења.

1.3. ПОЛАЗНЕ ХИПОТЕЗЕ

Главна хипотеза у раду гласи:

Развојем и имплементацијом модела паметне библиотеке заснованог на *big data* технологијама прилагођеног за примену у образовним институцијама могуће је унапредити образовни процес и побољшати свеукупне резултате учења.

На основу дефинисаног предмета истраживања и основне хипотезе, могу се издвојити следеће посебне хипотезе:

X1. Могуће је развити и применити модел паметне библиотеке заснован на *big data* технологијама.

X2. Примена *big data* система препоруке у паметним библиотекама доприноси унапређењу библиотечког пословања.

X3. *Big data* сервиси паметне библиотеке могу се интегрисати у образовни процес у циљу унапређења квалитета образовног процеса.

Даљим дефинисањем наведених посебних хипотеза, могу да се формулишу следеће појединачне хипотезе:

X1.1. Могуће је развити *big data* систем препоруке паметне библиотеке, посебно погодан за примену у образовним институцијама.

X1.2. Применом *big data* технологија и сервиса, могуће је интегрисати више извора различитих података у паметни библиотечки информациони систем.

X2.1. Могуће је реализовати евалуацију модела паметне библиотеке применом *big data* аналитике.

X2.2. Примена *big data* технологија у унапређењу информационе инфраструктуре образовне институције доприноси бољем пословном резултату.

X3.1. Корисници су задовољни квалитетом препорученог садржаја путем система препоруке паметне библиотеке.

X3.2. Психолошки, економски и технолошки аспекти су релевантни у процени перформанси модела паметне библиотеке применом *big data* технологија.

1.4. МЕТОДЕ ИСТРАЖИВАЊА

Од општенаучних метода коришћена је метода анализе и синтезе постојећих научних резултата, моделирање, аналитичко-дедуктивна и статистичка метода. Методама анализе и синтезе су анализиране теоријске основе паметних библиотека, *big data* технологија и система препоруке. Моделирање се користило за израду модела паметне библиотеке заснованог на *big data* технологијама, за моделирање *big data* инфраструктуре и сервиса. Аналитичко-дедуктивне методе су коришћене за анализу података о постојећим решењима, приступима, и о технологијама за развој софтверских компонента. Мерење релевантних параметара перформанси система, и анализа добијених резултата су обављени помоћу статистичких метода и *big data* аналитике.

У експерименталном делу рада је извршена евалуација развијеног модела. Добијени резултати су потврдили главну и посебне хипотезе истраживања.

Истраживање се може класификовати као мултидисциплинарно јер обухвата методологију, информатику, рачунарство, статистику и психологију. Резултати истраживања су представљени текстуално, описивањем и графички кроз више слика, дијаграма и табела са упоредним резултатима.

1.5. СТРУКТУРА И ОРГАНИЗАЦИЈА РАДА

Уводно поглавље описује предмет истраживања, циљеве истраживања, полазне хипотезе, методе истраживања, структуру и организацију рада.

У другом поглављу су анализирана постојећа решења и осврт на литературу библиотечког пословања традиционалних и дигиталних библиотека и система препорука у *big data* окружењу. У овом поглављу је такође анализиран утицај нових технологија у надоградњи библиотечке инфраструктуре и изградњи паметног библиотечког информационог система; као и идентификација начина за решавање изазова створених брзим информационим развојем и технолошким променама.

Треће поглавље описује технологије примењене за развој модела. Ово поглавље је подељено у два потпоглавља: *big data* и системи препорука. Прво потпоглавље описује *big data* концепт, инфраструктуру и сервисе. У другом потпоглављу су описани системи препорука. Приказана је реализација система и извршена је компаративна анализа применом *Mahout* оквира коришћењем скупа података евиденције сервера електронске продавнице књига.

Четврто поглавље обухвата развој предложеног модела паметне библиотеке заснованог на *big data* технологијама. Приказана је архитектура система и пројектовање *big data* решења. Развијен је модел паметне библиотеке заснован на *big data* технологијама и приказана је архитектура система која омогућава интеграцију више извора различитих података у *big data* екосистем који је посебно погодан за примену у образовним институцијама.

Примена развијеног модела, евалуација решења и анализа постигнутих резултата је извршена у петом поглављу.

Закључак обухвата преглед научних и стручних доприноса дисертације и друштвене импликације; планове за даља истраживања и приказ референтне литературе.

У прилогу је дат списак свих слика и табела, као и додатних слика.

2. АНАЛИЗА ПОСТОЈЕЋИХ РЕШЕЊА И ОСВРТ НА ЛИТЕРАТУРУ

Један од првих забележених покушаја прикупљања података је био средином трећег века пре нове ере када је *Pinakes* организовао каталог библиотеке наводећи имена аутора по абecedном реду (*Weinberger, 2007; Teets and Goldner, 2013*). Током времена, информације су се у каталозима библиотека селектовале и сортирале првенствено за физичку примену и физички свет (*Eliot and Rose, 2009*). Значајна иновација се догодила 2010. године када је Немачка национална библиотека објавила каталог библиотеке који повезује радове различитих аутора из истих области (*Svensson and Jahns, 2010*). Британска библиотека је 2011. године представила податке националне библиотеке као повезане податке, описујући модел ствари од интереса (енгл. *Things of Interest*) где је назив књиге повезан са људима, догађајима и местима (*Hodder, 2013; Teets and Goldner, 2013*). Према Изјави *Statement on Libraries and Development* коју је објавила Међународна федерација библиотекарских удружења и институција (*International Federation of Library Associations and Institutions – IFLA*): “Библиотека је једино место у многим заједницама у којима људи могу приступити информацијама које ће помоћи у побољшању њиховог образовања, развијању нових вештина, проналажењу послова, грађењу пословних пројеката, доношењу здравствених одлука и решавању питања животне средине.” (*IFLA, 2013; Cao et al., 2018*).

У библиотечком пословању, још од *Ranganathan*-ових *The Five Laws of Library Science*, поставља се питање примене нових технологија ради побољшања услуге корисницима и укупног доприноса науци (*Ranganathan, 1931*). *Toffler* (1971) је у својој књизи *Future Shock* поставио теорију да ће преоптерећење информација (енгл. *Information Overload*) довести до конфликта у доношењу одлука. Према ауторима: *Meier* (1963); *Lipowski* (1975); *Vollmann* (1991); *Hunt and Newman* (1997); *Grisé and Gallupe* (1999); *Wurman* (2001); *Hansen and Haas* (2001); *Eppler and Mengis* (2004); *Srinivasan* (2016), дата је класификација преоптерећења информација које је идентификовала академска заједница: прекомерно преоптерећење комуникација; преоптерећење информација добијених од сензора; когнитивно преоптерећење; преоптерећење знања у медицини; синдром информационог zasiћења; преоптерећење информација у оквиру сектора професионалних сервиса, од пословних консултација до пословних састанака и сл.

Кумулативно повећање обима података из различитих извора, релационих и нерелационих доводи до проблема пружања ефикасне библиотечке услуге корисницима. Библиотеке се суочавају са све већим изазовима како би биле у стању да задовоље све веће потребе својих корисника (*Cao et al., 2018*). Изазови се изражени у примарна три домена кроз: брзе промене у пројектовању и

реализацији нових информационих система; нове приступе анализи интензивних података; и разноврсности потреба корисника (*Cao et al.*, 2018).

Према Рубину (2017), образовне институције су у периоду транзиције када је у питању начин на који пружају своје библиотечке услуге. Нове методе доставе садржаја омогућавају електронским библиотекама да замене традиционалне препоруке библиотекара. Методе препорука треба да се прилагоде како би могле да одговоре на континуирано побољшавање нових технологија које нуде могућност нових облика испоруке информација. У складу са тим, природа и структура каталога образовних библиотека треба да се промени како би понудили садржај од већег интереса у складу са појавом нових економских модела (*Simović*, 2018). Ови захтевни услови чине паметну библиотеку неизоставним делом савременог библиотечког пословног система образовне институције, као и саставног дела континуираног процеса учења.

2.1. ТРАДИЦИОНАЛНЕ И ДИГИТАЛНЕ БИБЛИОТЕКЕ

Традиционално конфигурисани системи за складиштење и анализу података спречавају библиотеке да постигну конкурентску предност. Академске библиотеке су изгубиле монопол над пружањем информација јер су корисници прешли на онлајн научне изворе (*Chambers*, 2013). Традиционални библиотечки системи немају способност за управљање великом количином података, слика, аудио-видео и сензорских записа, као и других сложених типова података који се великом брзином генеришу из различитих извора. Структурирани системи база података који складиште велику већину информација нису прикладни за аналитичку обраду и суочени су са проблемима капацитета складишта (*Bekkerman et al.*, 2011). Вредни подаци у тим системима који могу носити важне информације, и који су сакупљени и ускладиштени по високој цени коштања, игноришу се и коначно бришу због ограниченог складишног простора (*Worlton*, 1971). Потешкоће се такође налазе у прикупљању, дељењу, визуализацији и дугорочном чувању података (*Ahrens et al.*, 2011; *Hey et al.*, 2009; *Horstmann and Brase*, 2016; *Bhat*, 2018). Осим тога, у традиционалним системима библиотечког пословања, препоруке дају библиотекари на основу њиховог искуства, физичких картица корисника и ускладиштених података (*Simović*, 2018). Библиотечки каталози треба да садрже довољно информација о библиотечкој грађи и корисничким преференцијама како би могли да одреде потенцијално идеалан резултат и да адекватно одговоре захтевима корисника (*Horstmann and Brase*, 2016).

За разлику од традиционалних, препоруке које се примењују у системима дигиталних библиотека, корисницима треба да помогну у проналажењу и одабиру потребних информација и извора заснованих на знању из велике количине

доступних података (*Porcel and Herrera-Viedma, 2010*). Аутори (*Balabanović and Shoham, 1997*) су представили дигиталну библиотеку *Fab*, Универзитета Станфорд, која примењује хибридни систем препоруке у комбинацији алгоритама базираних на колаборативном филтрирању и на садржају. *Mooney and Roy (2000)* су развили систем препоруке књига са алгоритмом машинског учења за категоризацију текста. *Naive Bayesian* класификатор текста је екстраховао информације како би направио карактеристике књига са циљем проналажења релевантног садржаја. *Renda and Straccia (2005)* су предложили персонализовану услугу дигиталне библиотеке под називом *CYCLADES*. *CYCLADES*-ов алгоритам се ослања на персонализоване информације корисника и организације у интегрисаном окружењу. На основу *Porcel*-овог истраживања 2009. године, који је развио хибридни систем препоруке за предлагање истраживачких ресурса универзитетских дигиталних библиотека (*UDL*) ради руковања флексибилним информацијама помоћу лингвистичких ознака стварањем релација са преференцијама корисника, *Serrano-Guerrero* је 2009. године представио систем препоруке који је интегрисао *GoogleWave* технологију са *UDL*-ом.

Док ови разматрани системи дигиталних библиотека углавном користе хибридне приступе који препоручују садржај комбинујући технике колаборативног филтрирања са системима базираних на садржају (*Lu et al., 2015*), савремени библиотечки информациони системи треба да пруже увид у податке користећи различите технике као што је *big data* аналитика из свих типова и извора података; треба да омогуће пословну аналитику и обраду уз повећање капацитета за достављање значајних информација и садржаја од интереса путем система препоруке. Такав систем треба да омогући свеобухватну логистичку и аналитичку платформу са опремљеним алатима и решењима који су у стању да задовоље потребе софистицираних и захтеве модерних, паметних библиотечких информационих система (*Simović, 2018*).

2.2. СИСТЕМИ ПРЕПОРУКА У *BIG DATA* ОКРУЖЕЊУ

Системи препорука применом *big data* технологија имају потенцијал да буду значајно унапређени интеграцијом вишеструких извора различитих података у паметни *big data* екосистем (*Simović, 2018*). Системи препорука налазе примену у различитим доменима пословања и имају заједничку особину која је препозната као важан извор информација које се корисницима нуде путем оцењивања садржаја и објављивања коментара према њиховом мишљењу (*Lin, 2014*). *Lee et al. (2015)* су предложили потребу за новим техникама *big data* аналитике ради смањења пристрасности у вредновању филмова, постављајући питања о поузданости оцењивања као непристрасног индикатора квалитета. Такође су открили да раније вредновање од стране онлајн заједнице као велике мреже људи

насупротив пријатеља (енгл. *Crowd versus Friends*) може имати различит утицај на будуће оцене корисника. *Gao et al.* (2018) су истраживали утицај културолошких фактора корисника приликом онлајн оцењивања хотела. Емпиријски су открили да корисници који долазе из земаља са великом удаљеношћу дају ниже оцене хотелима путем онлајн упитника.

У *big data* окружењу, системи препорука се широко примењују за предлагање контаката или активности на платформама друштвених мрежа (*Wu et al.*, 2014); да побољшају циљане огласе рекламне индустрије (*Вууа et al.*, 2016); као и за препознавање преференција корисника у контексту културних догађаја на основу анализе друштвених мрежа (*Milovanović et al.*, 2019). Практична примена система препоруке може развити нове маркетиншке стратегије интеграцијом тренутне ситуације корисника и њихове будуће потребе, нудећи контекстуално релевантне социјализоване препоруке (*Shen et al.*, 2013). *Lu et al.* (2015) су представили развој система препорука и груписали њихову примену у осам главних категорија: е-управа; е-пословање; е-трговина; е-библиотека; е-туризам; услуге е-ресурса; активности е-група; и е-образовање.

У е-образовању, од раних 2000-их, системи препорука постају све популарнији. На основу развоја традиционалних система е-образовања и више од петнаест година истраживачких студија на ову тему, развијена су многа практична и применљива решења система препорука за е-образовање (*Simović*, 2018). Ови системи имају за циљ да помогну студентима да изаберу курсеве и пронађу материјале за учење за које су заинтересовани. *Zaiane* (2002) је предложио изградњу софтверског агента који користи технике рударења података који је заснован на алгоритму правила придруживања за изградњу модела који представља понашање корисника. *Lu* (2004) је предложио персонализовану препоруку у е-образовању и представио систем који користи рачунарски модел за идентификацију преференција учења студената који се комбинује са правилима придруживања приликом генерисања препоруке. *Chen and Chao* (2008) су развили систем који придружује традиционалне наслове књига онлајн дискусионим форумима и друштвеним мрежама. На основу преференција корисника, чланови добијају поруке из онлајн заједнице која укључује линкове ка другим онлајн ресурсима. *Romero et al.* (2009) су предложили систем препоруке у е-образовању који користи технике онлајн рударења података како би препоручили линкове ка другим ресурсима у образовном систему за учење на даљину. *Sarriano et al.* (2014) су развили систем препоруке у е-образовању заснован на хибридном алгоритмима. Овај систем препоручује циљеве учења генерисањем препоруке кроз искуства и потребе студената током образовног циклуса. Хибридни алгоритам генерише препоруку кроз три корака: мапирање, процена корисности и циљеви вишег нивоа учења. Проширење е-образовног система предложили су *Santos et al.* (2014) путем пројектовања персонализованих препорука. Ова студија показује да је изградња

персонализованог система е-образовања процес који треба да одговори потребама студената током читавог животног циклуса е-учења. *De Meo et al.* (2017) су формирали класе е-образовања вредновањем поверења и вештина студената предлагањем модела који је имао за циљ успостављање и еволуцију часова за е-учење на темељу информација доступних са друштвених мрежа.

Неке од релевантних научно-истраживачких студија које су повезале системе препорука са *big data* аналитиком укључују (*Khan et al.*, 2017): пословну аналитику (*Duan and Xiong*, 2015); климатске промене (*Lu et al.*, 2011); аналитику клијената банкарских услуга (*Sun et al.*, 2014); паметне градове (*Khan et al.*, 2015); аналитику друштвених мрежа (*Burnap et al.*, 2014); здравствену аналитику (*Raghupathi and Raghupathi*, 2014); аналитику система управљања железницом (*Thaduri et al.*, 2015); и интелигентни транспорт (*Chandio et al.*, 2015) (*Simović*, 2018). Такође, *Saboo et al.* (2016) су предложили временски променљив модел ефеката за руковање сложеним повезаним подацима са *big data* аналитиком за реалокацију ресурса у маркетиншким стратегијама. *He et al.* (2014) су представили систем препоруке на друштвеним мрежама заснован на *Hadoop*-у са циљем генерисања препоруке базиране на сличности корисника на друштвеним мрежама ради проналажења пријатеља и садржаја. *Ismail and Al-Feel* (2015) су предложили *Hadoop* систем препоруке за научно-истраживачке радове. Исте године, *Yao et al.* (2015) су развили *big data* екосистем заснован на *Hadoop* сервисима ради брже обраде података у здравственим службама и клиничким истраживањима. *Wang* (2016) је пројектовао и реализовао мрежни систем препоруке заснован на *Hadoop* платформи. *Yi et al.* (2017) су представили методу препоруке библиотечке грађе која је заснована на асоцијативним правилима комбинованим са *Artificial bee colony* алгоритмом ради генерисања персонализованих препорука наслова књига користећи историјске податке о њиховим зајмовима у библиотеци. *Amato et al.* (2017) су представили систем мултимедијалних препорука за друштвене мреже назване *SOS. NoSQL* интеграција релационе базе података са *Hadoop* екосистемом је био пројекат *Rodrigues et al.* (2018). Циљ овог пројекта је био имплементација модела електронске трговине и система за управљање трансакционим подацима са паметних картица ради обраде великих скупова података применом различитих *big data* сервиса.

Важно је напоменути да је једна од примена *big data* аналитике која је мање истраживана, систем препоруке паметне библиотеке образовне институције. Имплементација различитих скупова података и њихових извора у реализацији модела паметне библиотеке заснованог на *big data* технологијама може имати користи, како за образовну институцију, тако и за потребе студената у континууму учења.

2.3. ПАМЕТНЕ БИБЛИОТЕКЕ

Као нови модел библиотеке, паметна библиотека заузима место представника нове идеје и праксе. Паметна библиотека не ствара само нову вредност која помаже корисницима у проналажењу персонализованог садржаја, већ побољшава библиотечке операције интеграцијом *big data* технологија (Liu and Shen, 2018). Примена нових технологија у надоградњи библиотечке инфраструктуре и реализацији паметног библиотечког информационог система идентификује начине за решавање изазова брзог информационог развоја и технолошких промена. Паметни библиотечки информациони системи треба да обухвате интеграцију великих количина штампаних, мултимедијалних и дигиталних материјала различитих типова и различитих ресурса. Примена система за аутоматизацију пословних процеса и употреба пословне интелигенције у стварању паметних окружења обезбеђује корисницима иновативни сервис за претрагу и коришћење садржаја (Zimmerman and Chang, 2018). Примера ради, 2004. године, библиотеке, музеји и универзитети у Отави, у Канади, основали су коалицију *паметних библиотека* (Wang, 2011) користећи јединствен претраживач за пружање онлајн услуга коришћења е-извора са једног места. Kwok et al. (2008) објашњавају да традиционалне библиотечке операције захтевају велике количине људског напора и времена за обављање задатака управљања инвентаром, наводећи да *RFID* технологије могу да унапреде и побољшају перформансе библиотечких пословних процеса, посебно у процесу регистрације књига, задуживања и контроле залиха. Shiwei (2012) је описао три главне функције паметне библиотеке кроз: интерконекцију различитих извора, интеграцију разноврсних података и персонализованост доступног садржаја.

У проширењу аутоматизованих библиотечких услуга, изградња паметне библиотеке обухвата интеграцију апликација паметних телефона у њихов систем користећи текст и функције електронске поште за комуникацију са корисницима и објављивање видео записа, како би корисницима на даљину омогућили приступ различитим услугама (Zimmerman and Chang, 2018). Brian et al. (2014) су применили *IoT* и мобилне технологије за управљање библиотечком грађом и комуникацијом са паметним телефоном корисника. Користили су локални систем позиционирања (енгл. *Local Positioning System – LPS*) и уграђене ознаке (енгл. *Near Field Communication – NFC tags*) на књигама да би обезбедили комуникацију сервера библиотеке са мобилним телефоном корисника. Примарни циљ је био да се пројектује и реализује систем који је у могућности да у реалном времену провери да ли је књига доступна у библиотеци, и уколико јесте, да корисника усмери ка њеној тачној локацији на полици. Jerkov et al. (2015) су сматрали да је паметна библиотека *брокер знања* (енгл. *Knowledge Broker*) који проналази квалификоване изворе критичког знања за решавање специфичних проблема (Zimmerman and Chang, 2018). Представили су инфраструктуру која поред

библиотечког простора и места, повезује паметне библиотеке са паметним градовима.

Међутим, постоји мало теоријског истраживања о паметним библиотекама – шта представља паметну библиотеку и како се она разликује од других типова библиотека (Cao et al., 2018). Концептуализација паметне библиотеке може да помогне истраживачима и практичарима да постигну консензус о томе како побољшати пословање постојећих библиотека; како успоставити основне компоненте паметне библиотеке; и како одговорити захтевним технолошким изазовима и корисничким потребама. Темељна концептуализација омогућава будућа истраживања у развоју паметних библиотечких информационих система. Постоји низ концептуалних варијанти насталих заменом термина паметно са другим терминима као што су дигитално, интелигентно и комбиновано (Cao et al., 2018).

Паметна библиотека треба да буде усмерена и фокусирана на кориснике и прилагођена њиховим потребама. Да би се то постигло, термин паметно у датом контексту значи да библиотека треба да буде у стању да по аутоматизму препозна потребе корисника и да обезбеди ресурсе и услуге како би се те потребе задовољиле. Са друге стране, термин паметно у технолошком контексту подразумева аутоматске рачунарске принципе, као што су самоконфигурација, самозаштита и самооптимизација (Spangler et al., 2010).

У пракси, библиотека *Mansotto* Универзитета у Чикагу, користила је паметну технологију за успостављање роботског подземног система за управљање библиотечком грађом. Примењена технологија је заснована на аутоматизованим роботима који су били у стању да управљају складиштем од 3,5 милиона књига у 1/7 од укупног простора потребног у конвенционалним складиштима што је резултовало великом уштедом простора за складиштење и обезбедило аутоматизацију процеса рада (Thompson and Works, 2012; Kalee, 2012; Cao et al., 2018).

Нове технологије као што су *big data*, рачунарство у облаку и *IoT*, примењују се у свакодневном животу људи и разматрају се у више научних дисциплина (Nolin and Olson, 2016). Паметне куће, паметне зграде и друга паметна окружења као што су аеродроми, болнице или универзитетски кампови, опремљени су мноштвом мобилних терминала, уграђених уређаја, повезаних сензора и актуатора (Klein and Kaefler, 2008). Због тога се може извести закључак да се постојећа паметна технологија може применити у изградњи паметних библиотека и стварања система који би био у могућности да у потпуности сагледа потребе корисника, омогући паметно управљање и побољша ефикасност рада.

У табели 1 су приказане дефиниције паметне библиотеке идентификоване у научно-истраживачкој литератури (Cao et al., 2018). Ове дефиниције се фокусирају на технологију изградње паметне библиотеке и полазна су тачка за истраживања која могу бити водичи ка њиховој реализацији и смернице ка будућем унапређењу постојећих решења.

Табела 1. Дефиниције паметне библиотеке

Научне студије	Дефиниција
Aittola et al., 2003	Паметна библиотека пружа услуге мобилне библиотеке која није ограничена простором, и која помаже корисницима да пронађу књиге и сродне информације.
Miller et al., 2004	Паметна библиотека примењује апликације квалитетне инжењерске праксе са циљем смањивања вероватноће грешака и повећања корисничке способности за идентификовање и исправљање грешака, када и уколико се догоде.
Yan, 2010	Паметна библиотека је модел паметнијег начина интеракције корисника и библиотечких система користећи нову генерацију информационих технологија, како би се побољшала јасност, флексибилност и прилагодљивост паметних услуга и управљања.
Wang, 2011	Паметна библиотека реализује асоцијације између књига; књига и људи; људи и књига у било које време и на било којем месту. Дигитализација, умрежавање и интелигенција су информационе и техничке основе паметне библиотеке. Паметна библиотека је оријентисана према људима и одрживим развојем, са циљем задовољавања све већих потреба читалаца за информацијама.
En, 2012	Паметна библиотека је напреднији развој хибридне библиотеке и дигиталне библиотеке. У <i>IoT</i> окружењу, паметна библиотека се ослања на технологију рачунарства у облаку и интелигентне инфраструктуре; остварује асоцијације између књига, књига и људи, и људи и људи, пружајући интелигентне услуге корисницима.
Li and Dong, 2016; Cao et al., 2018	Паметна библиотека се односи на <i>висprenost</i> зграде библиотеке кроз интеграцију опреме за изградњу, рачунарске мреже, комуникационе технологије и надзор сензора.

Научна истраживања модела паметних библиотека се заснивају на перспективи једне димензије, као што је техничка димензија (Yan, 2010; Wei, 2012); димензија услуга (Xiaoxia et al., 2011; Chen and Zhang, 2016); димензија образаца/узорака

(En, 2012); или да се заснивају на вишедимензионалном приступу (Wang, 2011). Примера ради, библиотека *Chongqing* Пекиншког Универзитета је заснована на перспективи система (Xiaoxia et al., 2011), док је *Taipei* библиотека заснована на перспективи сервиса (Liu, 2015; Cao et al., 2018). Библиотека Универзитета *Nanjing* се заснива на техничкој перспективи, али игнорише корисничко искуство (Kuilin and Bo, 2015) које је склоно континуираним и динамичким променама током времена коришћења система.

У складу са наведеним, потребно је разматрање обједињених концепата система паметних библиотека приликом изградње скупа заједничких вишедимензионалних компоненти. На основу анализе научне-истраживачке литературе о паметним библиотекама и примера практичних решења, као што наведено, различити термини се могу категоризовати у три димензије: технологија, сервис/услуге, и корисници (приказано у табели 2) (Cao et al., 2018). Ове концептуалне варијанте су међусобно повезане са значајним преклапањем дефиниција и комплексности њихове примене.

Табела 2. Категоризација концептуалних релација димензија и дефиниција паметне библиотеке

Димензија	Концепт	Дефиниција	Научне студије
Технологија	Дигитална библиотека	Дигитална библиотека је организација која пружа ресурсе. Њена сврха је сакупљање, изградња, приступ, откривање, пренос и одржавање великог броја дигиталних радова; и да обезбеди издржљивост колекције током времена тако да се може лако и економично користити.	<i>Nikolaidou et al.</i> , 2005; <i>Kovacevic et al.</i> , 2010; <i>Diekema</i> , 2012; <i>Lagzian et al.</i> , 2013
	Интелигентна библиотека	Интелигентни библиотечки системи омогућавају корисницима да преузму информације на основу онога што мисле, а не на основу онога што кажу. Омогућавају библиотекама да постану интерактивније, тачније и корисније.	<i>Bailey</i> , 1991; <i>Wahono</i> , 2000; <i>Dent</i> , 2007
	Хибридна библиотека	Применом различитих технологија у изградњи библиотечких информационих система, хибридна библиотека комбинује предности традиционалних и дигиталних библиотека; електронски и штампани ресурси коегзистирају, стварајући сложено окружење које корисницима пружа услуге.	<i>Rusbridge</i> , 1998; <i>Pinfield et al.</i> , 1998; <i>Cao et al.</i> , 2018
Сервиси / Услуге	Комбинована библиотека	Комбинована библиотека је облик библиотеке засноване на Интернету и новим технологијама. Са фокусом усмереним	<i>Bell and Shank</i> , 2004; <i>Sinclair</i> , 2009; <i>Bogorin et al.</i> , 2013

	<p>Свеприсутна (енгл. <i>Ubiquitous</i>) библиотека</p>	<p>на корисничке потребе, комбинована библиотека омогућава услуге управљања ресурсима података и ситуационих знања. Свеприсутна библиотека је дигитална библиотека где корисници могу путем информационих уређаја приступити информацијама које им требају било где и било када. Циљ свеприсутне библиотеке је да корисницима обезбеди интерактивне информативне ресурсе и услуге.</p>	<p><i>Kaske, 2004; Bae et al., 2007; Cao et al., 2018</i></p>
<p>Корисници</p>	<p>Мобилна библиотека</p>	<p>Мобилна библиотека се односи на услугу која омогућава приступ библиотечким ресурсима помоћу паметних телефона и друге опреме за мобилни приступ. Пројектовање мобилне библиотеке се ослања на мобилне платформе које су оријентисане према захтевима корисника. Одражава принцип који је фокусиран на кориснике у процесу пројектовања библиотечких сервиса.</p>	<p><i>Huang and Wang, 2011; Wei and Yang, 2017</i></p>

Напредна технологија је основа и премиса паметне библиотеке коју чине *big data*, технологија интелигентне обраде података, рачунарство у облаку, виртуализација, рударење података, вештачка интелигенција и *IoT* (Cao et al., 2018). Tim Berners-Lee је 2009. године предложио први корак повезивања података (енгл. *Linked Data*) у облику у којем их машине могу природно разумети или претворити у разумљиву форму (Bizer et al., 2009). Библиотечки информациони системи примењују технике рударења података како би помогли и утицали у доношењу одлука (Cullen, 2005). Cox and Jantti (2012), су придружили библиотечке податке Универзитета Wollongong са демографским и академским подацима како би приказали корелацију коришћења библиотечке грађе. Библиотечко рударење података у изградњи система препорука за библиотечке материјале резултат је истраживања Chen and Chen, 2007; Tsai and Chen, 2008; Kovacevic et al., 2010. Методе рударења података могу произвести појединачне и персоналне услуге корисницима тако што пружају одговарајуће информације и асистирају у доношењу одлука. Осим тога, методе рударења могу повезати различите наслове и садржаје стварајући богатије индивидуалне ресурсе за кориснике (Renaud et al., 2016; Cao et al., 2018).

Са друге стране, *IoT* се широко примењује у здравству; ланцу снабдевања; логистици; транспорту; анализи могућих пожара; аутоматизацији зграда и интелигентних кућа; паметним мрежама; паметним градовима; и управљању енергијом (Fang et al., 2014; Xu et al., 2014; Wang et al., 2014; Da Xu et al., 2014; Rashid et al., 2016; Chen, 2017; Civerchia et al., 2017; Kim, 2017; Lai et al., 2017; Liu et al., 2017; Bi et al., 2018; Xu et al., 2018; Wang et al., 2018; Davidovic et al., 2018). *IoT* апликације су се из домена привредног сектора и здравства прошириле и на образовање. Факултети и универзитети активно граде паметна окружења заснована на *IoT* технологији како би осигурали интеграцију и дељење дигиталних ресурса, управљање библиотечком, окружењем у настави, наставним инструментима и опремом, укључујући идентификацију наставника и студената (Wang, 2014; Zhao, 2013; Zong et al., 2014). Wójcik (2016) наводи да би *IoT* технологија могла имати потенцијал да се користи у библиотечким услугама слично како се примјењује у комерцијалном сектору и да ће донети низ значајних промена за библиотеке (Du and Liu, 2014; Sun, 2014), претварајући их у паметне (Xu, 2014). Li (2013) истиче да се многе библиотечке услуге могу реализовати применом *IoT* технологија, стварајући инвентар паметним. Сходно томе, е-ресурси и дигиталне библиотеке постају све важнији канал за добијање тражених информација (Hu and Zhang, 2016). Примена *IoT* технологија у библиотечком пословању је приказана у многим научним истраживањима. Andersen је 2002. године дефинисао однос између комуникационих технологија и концепта организације знања у библиотекама. Liu and Sheng (2011) су истражили развојни правац Интернет ствари у области управљања библиотекама и програма промоције. Ma et al. (2011) су представили интегрисани систем управљања са вишеслојном архитектуром заснованом на *IoT*-у за управљање студијским просторијама у библиотекама. Sun et al. (2014) су применили *IoT* технологију за откривање, идентификацију и анализу понашања корисника у циљу побољшања библиотечке услуге, управљања ресурсима и постизања оптималне

конфигурације. *Yao and Song* (2014) су предложили *IoT* интелигентни систем контроле за уштеду енергије светла у библиотекама. *Wei et al.* (2014) разматрају модел интелигентног библиотечног система за управљање информацијама заснованог на рачунарству у облаку и *IoT* технологији. *Ma* (2015) истиче да је информационо-комуникациона инфраструктура заснована на *IoT*-у важна за библиотечке информационе системе и ефикасан метод за побољшање рада библиотеке. *Cheng et al.* (2016) представљају клијентски софтвер за мобилне уређаје како би се побољшала ефикасност претраживања књига, управљање ресурсима, и сачувала материјална средства.

Као корак даље, *Cox et al.* (2018) наводе неизбежност утицаја вештачке интелигенције (енгл. *Artificial Intelligence – AI*) у будућој изградњи паметних библиотечких система. *Fernandez* (2016) проширује примену вештачке интелигенције у библиотечким системима постављајући питање, на које системе уопште неће утицати *AI* технологија. *Arlitsch and Newell* (2017) наводе да ће *AI* трансформисати библиотечке услуге заувек, континуирано модификовати потребне вештине и алате за опслуживање корисника, истовремено мењајући динамику развоја наших заједница.

Ипак, истраживачи и научници потврђују да образац промена остаје нејасан – ставови, разумевање и очекивања у академском библиотечком сектору су тумачени на различите начине (*Cox et al.*, 2018). То је правац који повезује аутоматизацију процеса, Интернет ствари, обраду података, роботiku, и подршку у одлучивању. *Smith* (2016) дефинише оквир вештачке интелигенције који се састоји од: *big data* технологија, машинског учења, аналитике и визуализације. Вештачка интелигенција се гради на *big data* технологијама (*Cox et al.*, 2018) јер комбинација великих скупова различитих података са рачунарском снагом ствара потенцијал за нове *AI* нивое. Специфичне области развоја се односе на способност паметних система да побољшају обраду података током свог рада независно од експлицитног програмирања.

Као што је приказано, анализа научно-истраживачке литературе указује на недостатак истраживања у развоју модела паметне библиотеке заснованог на *big data* технологијама који омогућава аутоматизацију пословних процеса применом система препоруке који је посебно погодан за примену у образовним институцијама. Остваривање ових захтевних услова омогућава библиотекама да достигну конкурентску предност, друштвену одговорност и да остваре своју примарну друштвену улогу.

3. ТЕХНОЛОГИЈЕ ПРИМЕЊЕНЕ ЗА РАЗВОЈ МОДЕЛА

3.1. *BIG DATA*

International Data Corporation – IDC је у својој студији проценила да ће укупан раст доступних информација 2020. године износити 40ZB и предвиђа да ће тржиште *big data* технологија расти по годишњој стопи од 23,1% (*Edjlali et al.*, 2012; *Grover et al.*, 2018). На брзину генерисања различитих података у јединици времена утичу истраживачко-научне институције, приватни и државни сектор, интеграцијом све више уређаја у своје пословне процесе (*Boyd and Crawford*, 2012; *Cukier and Mayer-Schoenberger*, 2013; *Panetta*, 2016).

Традиционално, подаци су сакупљени и централизовани у складишту података организације на једном серверу и изграђени су на *SMP* технологији (енгл. *Symmetric Multi Processing*). Са *SMP* технологијом, проширивање капацитета би значило додавање јачег хардвера за складиштење и затим враћати податке из претходне базе. То је било неопходно јер би база премашивала дозвољени капацитет јер није скалабилна, а нема начина за накнадно додавање инкременталне снаге процесора или омогућавање синхронизације кеша језгара више процесора (*Microsoft*, 2016). Када су количине података веће, њихове комплексне релације захтевају вишефазну обраду (*Pal*, 2016). Традиционални системи за управљање подацима су пројектовани тако да складиште историјске информације и корисничке радње кроз евиденције сервера са претпоставком да ће ти подаци бити сачувани сада, а анализирани касније. Ипак, како брзина ускладиштених података расте, корист од њихове обраде се вишеструко повећава (*Microsoft*, 2016).

Како је све више пословних активности дигитализовано, организације користе податке у реалном времену да промене, изграде или оптимизују своје пословање, као и да обаве финансијске трансакције и ангажују се у динамичким догађајима као што је берзанско пословање (*Andersen et al.*, 2003). Традиционални системи нису направљени да подрже њихову обраду у реалном времену што доводи до смањења перформанси и краћим временом остваривања вредности. Традиционални приступ складиштењу и анализи није пројектован да управља великом количином, типовима и брзином генерисања података (*Lin and Dyer*, 2010).

Изван традиционалних извора трансакционих система пословних софтвера и њихових апликација, нови типови података опредељују организације за нови приступ њиховој анализи, обради, складиштењу и откривању увида у знање из информација које подаци носе. Такође, опадање трошкова елемената рачунарства: складиштења, меморије, пропусног опсега мреже и сл, значи да ће интензиван скуп разноврсности података постати економски исплатив (*Hashem et al.*, 2015). Комбинација нових извора и типова података са приступачном опремом отварају ново доба електронског пословања (енгл.

Data Science, eScience) у којој постоји велика количина дигиталних информација које се тичу било које теме од пословног интереса. Е-продавнице, друштвене мреже, паметни телефони, Интернет ствари са подршком за даљински надзор, подаци са *GPS* уређаја за одређивање локације, е-трансакције и остале електронске комуникације производе велике количине података својим дневним, уобичајеним активностима (*Microsoft*, 2016). Ови разноврсни типови података се не могу лако уклопити у пословне операције компанија и не могу бити исплативи за *ETL* процесе (енгл. *Extracted, Transformed, Loaded*) у релационим системима за управљање подацима (*Golfarelli and Rizzi*, 2009). Масивни скупови разноврсних података сложених структура представљају потешкоће за складиштење, анализу и визуализацију резултата за даље пословне активности. Доступни подаци су често неструктурирани, неорганизовани у базама и велики, носећи информације од значаја које чекају да буду анализирани и употребљени. Компаније реагују на растуће неструктуриране податке што их приморава да усвоје и савладају знања нових технологија и да адаптирају своје информационе системе са новом инфраструктуром погодном за *big data* аналитику (*Sagiroglu and Sinanc*, 2013).

Модерни информациони системи треба да омогуће управљање, обраду и анализу великих скупова података различитих типова са много већом могућношћу испоруке информација од значаја у реалном времену. *Big data* аналитика је донела строге услове и технике за обраду података, по аутоматизму стварајући *big data* парадигму незаобилазним делом пословних процеса.

Big data дефинише пројектовање и реализацију скалабилне и толерантне на грешке (енгл. *Fault Tolerant*) инфраструктуре система способног за паралелно процесирање, складиштење и визуализацију великих количина различитих података који се стварају великом брзином из различитих извора (*Simović*, 2018).

3.1.1. *BIG DATA* КОНЦЕПТИ

Laney (2001) је дефинисао основне изазове које стварају велике количине доступних података кроз три димензије (енгл. *Volume, Velocity, Variety*):

- *Volume* – како укупну величину и акумулативну запремину скупа података који је изван капацитета постојећег система за управљање релационим базама података, обрадити;
- *Velocity* – како ускладиштити податке који се великом брзином генеришу кроз динамичку интеракцију корисника; и
- *Variety* – како анализирати различите формате некомпатибилних и неконзистентних структура података.

Гартнер је 2012. године проширио *big data* дефиницију и објавио да информациони системи захтевају нове облике обраде како би омогућили боље доношење одлука,

откривање увида у знање и оптимизацију пословних процеса (Beyer and Laney, 2012). IBM је додао још један *big data* атрибут који се односи на истинитост података (енгл. Veracity) (Erl et al., 2016). Zikopoulos and Eaton (2011) су објаснили разлог иза новог *big data* атрибута представљајући димензију тачности као одговор на квалитет извора података у суочењу са *big data* иницијативама. Према научним истраживањима, постоји више од десет *big data* атрибута *big data* концепата (Markus, 2015).



Слика 1. Комплексност *big data* концепта у модерном електронском пословању (адаптирано према: Microsoft, 2016)

Big data концепти су базирани на доминацији нових модела информационих структура, нових извора података, и на могућношћу интеграције нерелационих и неструктурираних података из екстерних извора са интерним пословним аналитичким системима (приказано на слици 1). Примена *big data* концепта може остварити вредност у широком спектру развоја нових модела *big data* инфраструктура како нови извори података мењају начин извршавања пословних процеса. Примери нових модела *big data* концепата укључују оптимизацију информационе инфраструктуре, анализу друштвених мрежа, оптимизацију протока саобраћаја, оптимизацију веб апликација, интеграцију информација о локацији, анализу флукуације, природно истраживање ресурса, временску прогнозу, здравствену заштиту, откривање превара, истраживање природних кретања, анализу оглашавања, мониторинг у реалном времену и др (Microsoft, 2016).

3.1.2. *BIG DATA* ИНФРАСТРУКТУРА

Са великим и брзим растом доступних информација, кастомизована инфраструктура *Google* сервиса је достигла границе скалабилности (*Olson, 2010*). Две компоненте су биле кључне у пројектовању нове инфраструктуре за процесирање података (*Dean and Ghemawat, 2008*):

- *Google* систем датотека (енгл. *GFS – Google File System*) – поуздано и скалабилно складиште генерисаних података са обезбеђивањем толерантности на грешке (енгл. *Fault Tolerant*); и
- *MapReduce* – систем за процесирање података који омогућава дељење и паралелно обављање задатака на великом броју сервера.

Doug Cutting реализује пројекат отвореног кода *Apache Hadoop*, заснован на инфраструктури *Google*-а, који се такође састоји од две кључне компоненте (*Olson, 2010*):

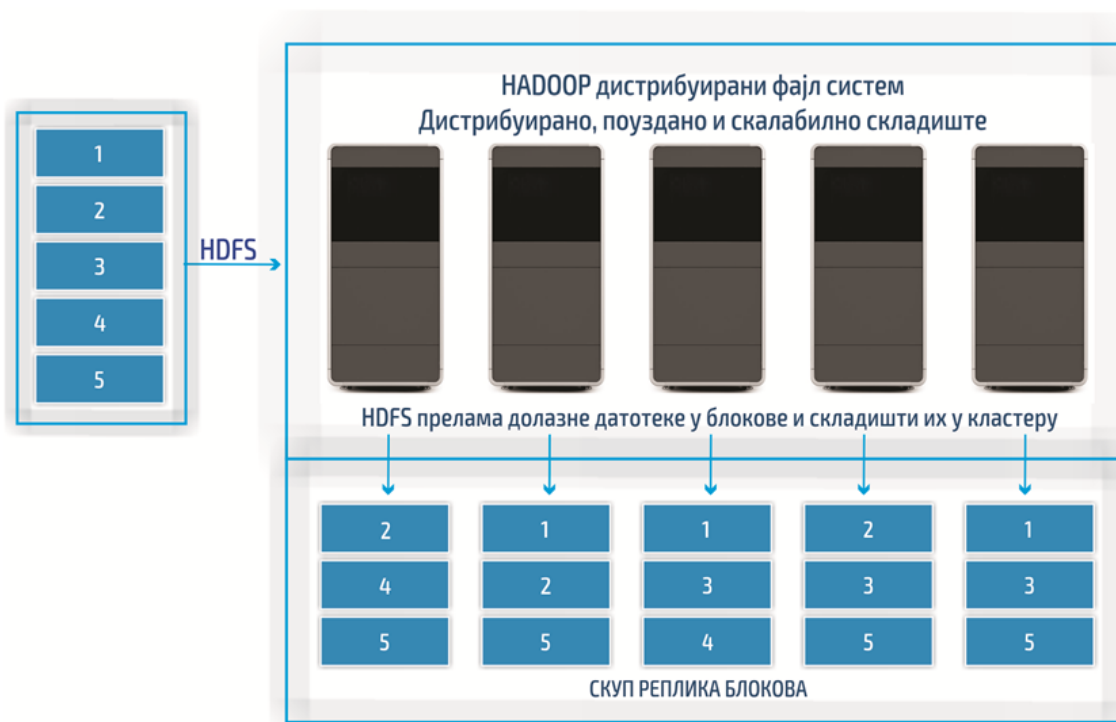
- дистрибуирано складиште (енгл. *Hadoop Distributed File System – HDFS*), и
- дистрибуирани систем за обраду заснован на *Google* решењу за паралелно процесирање података (енгл. *Distributed Processing System – MapReduce*).

Подаци се чувају на *HDFS*-у који обезбеђује скалабилно, прилагодљиво и отпорно на грешке складиште. Приликом складиштења, датотеке се разграђују по блоковима величине од 64MB или 128MB, и сваки блок је копиран на више од једног сервера (обично 3) у кластеру (*White, 2012*).

Ова репликација омогућава (*Olson, 2010*):

- отпорност и толерантност на грешке – где губитак једног диска, рачунара или сервера неће уништити датотеку, и
- перформансе – где било који блок може да се прочита са једног или више сервера притом побољшавајући проток података (енгл. *Data Flow*).

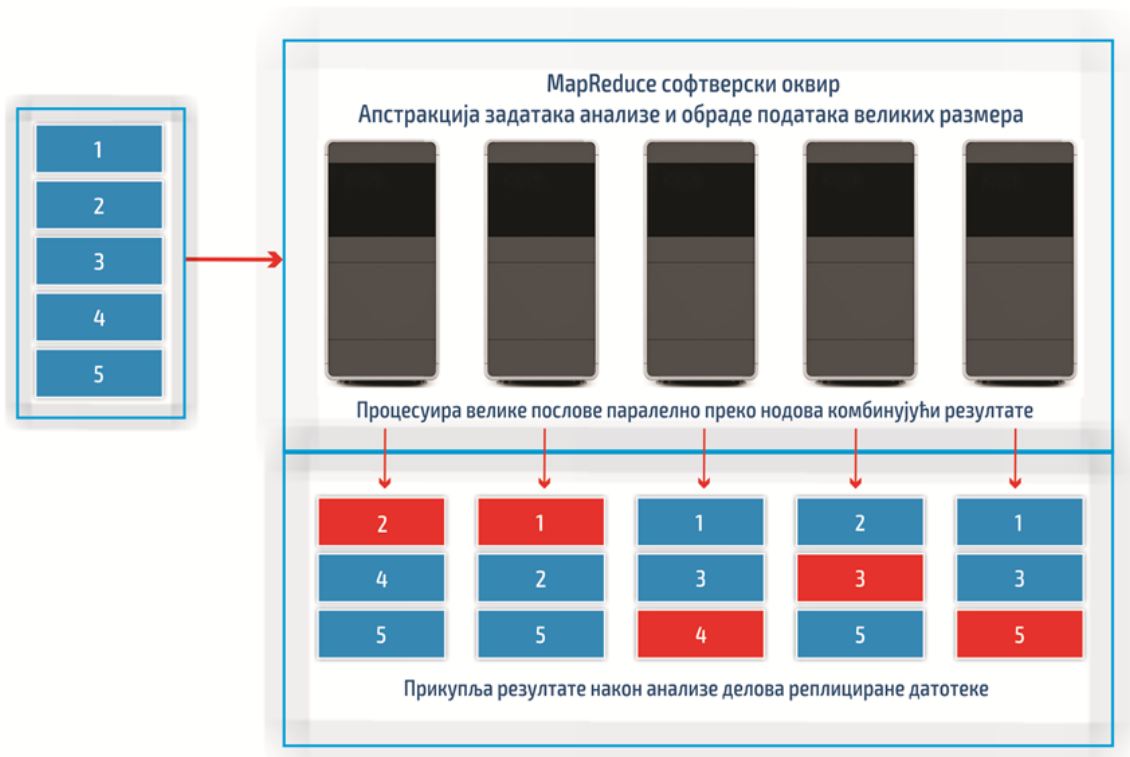
HDFS ствара више реплика сваког блока података и дистрибуира их на друге чворове широм кластера омогућавајући поуздан и брз приступ (приказано на слици 2).



Слика 2. Инфраструктура дистрибуираног складишта – HDFS (адаптирано према: Olson, 2010)

HDFS обезбеђује доступност података тако што континуирано прати рад сервера у кластеру и блоковима којима сервери управљају. Индивидуални блокови подлежу провери и контроли рада. Када се блок прочита, утврђује се исправност (да ли је вредност која је забележена, исправна). Уколико је блок оштећен, бива обновљен са једним од његових реплика, а уколико сервер или диск откаже, сви подаци које складишти реплицирају се на други чвор или на друге чворове у кластеру из читаве колекције реплика (Olson, 2010). Као резултат, *HDFS* детектује и компензује грешке на серверу или проблеме на дисковима на другим чворовима у кластеру обезбеђујући поуздано складиште.

Сама *HDFS* инфраструктура није довољна да би се створио одговарајући ниво адаптације инфраструктуре *big data* екосистема. Обзиром да је експоненцијални раст укупне количине података довео до многих виталних изазова у електронском пословању, постојећи системи су постали неадекватни за њихову обраду (Olson, 2010). Да би то превазишао, *Google* је представио нови модел програмирања – *MapReduce* (приказано на слици 3), другу главну компоненту за паралелно и дистрибуирано процесирање података (Dean and Ghemawat, 2008).

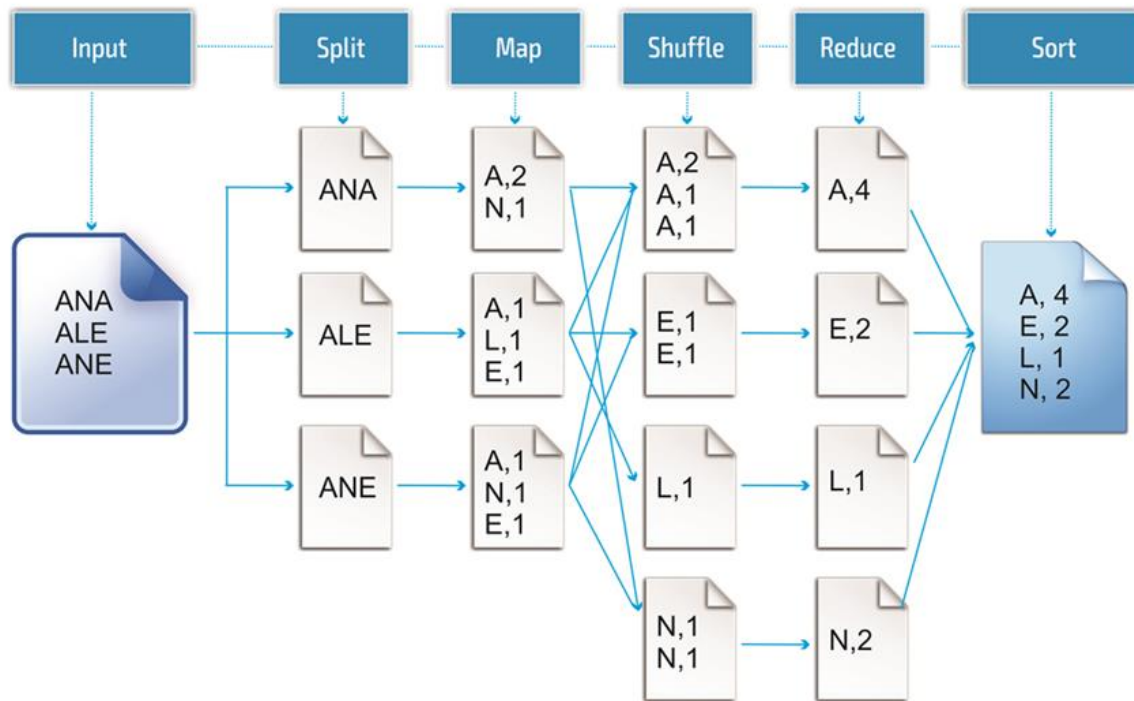


Слика 3. Инфраструктура MapReduce софтверског оквира (адаптирано према: Olson, 2010)

MapReduce експлоатише дистрибуирану архитектуру складишта ради прилагодљивог, скалабилног, поузданог и паралелно-оријентисаног алгоритма за процесирање датих задатака. Ова комбинација је консолидација свих типова података и поуздана платформа за чување података која испоручује дистрибуирано извршавање (Olson, 2010; Lin and Dyer, 2010; White, 2012).

Улазни подаци за MapReduce процесирање се обично преузимају од датотека које су уčitане у HDFS. Датотеке из HDFS се равномерно распоређују на све чворове у кластеру (Wang et al., 2013). MapReduce обухвата апстракцију која се зове улазна подела (енгл. *Input Split*) и која омогућава да сваки блок буде подељен на индивидуалне, појединачне записе (приказано на слици 4). Пре него што почне извршавање задатог посла, раздвајање улазних података за конфигурисани број задатака се израчунава како би се постигла подједнака дистрибуција радног оптерећења. У фази мапирања, улазна датотека је подељена на независне улазне поделе (енгл. *Input Split*) где свака подела садржи појединачни задатак мапера (*Map*) који описује MapReduce посао. Задаци мапирања се затим додељују чворовима у кластеру на основу физичке локације улазних датотека. Неколико задатака мапирања се могу доделити појединачном чвору који настоји да изврши што више задатака паралелно (Olson, 2010; White, 2012; Wang et al., 2013).

На крају фазе посла мапера (*Map*) резултати се сакупљају и редуцирају (*Reduce*). *MapReduce* процес осигурава да подаци буду сортирани и достављени редуктору тако да се излази свих мапера прикупљају и прослеђују процесу сортирања (енгл. *Shuffle and Sort*). Сортиран излаз се затим преноси на део редуције за обраду. Резултати су уписују назад у *HDFS* (*White, 2012*).



Слика 4. MapReduce – Input Split / Shuffle and Sort

Пројектовање и реализацију *big data* инфраструктуре прате три карактеристике (*Olson, 2010*):

- Велики системи за управљање подацима су другачије изграђени данас. Уместо великих, централизованих сервера, организације покрећу колекције уградних складишта на серверима, од којих сваки има више процесора и дискове капацитета изражених у ТВ. Старије, централизоване и дељене апликације споро и слабо раде у таквим условима.
- Разноврсност и комплексност расположивих података је брзо порасла. Управљање релационим базама података и структурирани системи немају могућност обраде великих и комплексних типова података. Ова комплексност је и квалитативне и квантитативне природе. Комплексни подаци не могу бити лако искоришћени у базама структурираних редова и колона. Подаци ове врсте су велики, а брзина којом се генеришу расте. Нове технологије за обраду података треба да буду способне за издвајање вредности скривених у великим и различитим подацима.

- *Big data* инфраструктура треба да омогући аналитику комбиновањем структурираних и неструктурираних података, и да обезбеди увид у информације из различитих извора.

Примена *big data* технологија у реализацији информационе инфраструктуре има доказане вредности у разним областима пословања. Банке и осигуравајућа друштва користе алгоритме за праћење понашања клијената на основу историје трансакција ради откривања сумњивих активности; компаније које се баве електронском трговином имају могућност да побољшају тачност својих система за препоручивање производа комбинујући трансакционе податке са евиденцијама активности корисника генерисаним на веб серверу или другим мрежним уређајима.

У енергетском сектору, компанија *Tennessee Valley Authority – TVA* је изградила систем прикупљања и анализе података под називом *Open Phasor Data Collector – OpenPDC* који је заснован на *big data* инфраструктури. *OpenPDC* прати количину долазећих података од сензора везаних за систем производње електричне енергије. Ови сензори, распоређени око електричне мреже, прате и извештавају функционалност и статус сваког генератора. Пажљиво праћење и брз одговор на могуће промене на мрежи омогућава екосистему да смањи или спречи кварове, као и да боље управља расположивим капацитетима. Инфраструктура *big data* екосистема *OpenPDC* је била у стању да смањи оперативне трошкове компаније и да обезбеди друштвено одговорно пословање са циљем боље контроле стаклене баште, емисионих гасова и других еколошких ризика (*Olson, 2010; Yu et al., 2013*).

Компаније које развијају *Hadoop* сервисе и алате и значајни његови дистрибутери на тржишту које нуде и услуге техничке подршке кроз надоградњу *Hadoop* модула су: *Hortonworks, Cloudera, Amazon Elastic MR, Microsoft, IBM, Pentaho, Fico, Jaspersoft, Apache Bigtop, Cascading* и други.

3.1.3. **BIG DATA ИНФРАСТРУКТУРА MICROSOFT**

Microsoft је са *SQL* сервером 2016 обухватио *big data* решења инфраструктуре за локално и *Cloud* окружење како би омогућио прикупљање и обраду релационих и нерелационих података. *Cloud* окружење као скуп мрежних сервиса пружа скалабилне, персонализоване, рачуарске инфраструктуре на захтев (енгл. *On Demand*), којима се може приступити на једноставан и продоран начин (*Ranjan et al., 2010; Buyya and Ranjan, 2010*). Концептуално, корисници захтевају рачуарску платформу или информациону инфраструктуру за извршавање израчунавања или покретања апликација унутар њих. *Cloud* омогућава корисницима услуге приступа хардверу, софтверу, ресурсима података, и интегрисаној платформи као сервису (енгл. *Platform as a Service*). *MapReduce* парадигма и његова инфраструктура је препозната као репрезентативна техника која омогућава технологију рачунарства у облаку (енгл. *Cloud Computing*) (*Wang et al., 2013*).

Big data инфраструктура *Microsoft*-овог серверског решења 2016 кроз интеграцију са *Cloud* окружењем омогућава да се избором правог распоређивања оптерећења могу доделити задаци и улоге у систему где једна величина, било да је простор или датотека, може одговарати дефинисаним нивоима приступа. Обухвата пројектовање вишеструких складишта – преко израде апликација које захтевају велике количине разноврсних података (структурираних, полуструктурираних и неструктурираних) до различитих нивоа приступа корисника који те апликације могу истовремено да користе. Инфраструктуру карактерише (*Microsoft*, 2016):

- *PolyBase* технологија која обезбеђује добијање траженог одговора на основу једног *T-SQL* упита (енгл. *Transact-SQL*) преко релационих и нерелационих база података, *SQL* сервера и *Apache Hadoop* оквира (*DeWitt et al.*, 2013);
- Подршка за решења складиштења структурираних и неструктурираних података, симетрично мултипроцесирање (енгл. *Symmetric Multiprocessing – SMP*), и обимно паралелно процесирање (енгл. *Massively Parallel Processing – MPP*) (*Fan and Huang*, 2013; *Hashem et al.*, 2015);
- Хибридно складиштење података са *Stretch DB* модулом;
- Подршка за повезивање више извора података;
- Подршка за проток неструктурираних података;
- *JSON* подршка (енгл. *Java Script Object Notation*);
- Унапређени примарни сервис (енгл. *Master Data Services – MDS*);
- Побољшање оперативних перформанси применом индекса заснованих на колони (енгл. *Columnstore indexes*);
- *SQL* серверски сервис за анализу, са ефикасним паралелним процесирањем и управљањем меморијским подацима (енгл. *SQL Server Analysis Services – SSAS*);
- Интеграција *R* језика која омогућава анализу података унутар локалне, релационе базе података;
- Изградња серверских развојних алата за: моделе применљиве у пословној интелигенцији (енгл. *Business Intelligence – BI Models*); моделе анализе (енгл. *Analysis Services Models*); и серверске моделе извештавања (енгл. *SQL Server Reporting Services – SSRS*);
- Способност публикавања извештаја оптимизованих за мобилне уређаје (енгл. *Mobile Report Publisher*).

PolyBase технологију карактерише могућност повезивања традиционалног складишта података са *Hadoop* окружењем и постављање упита над релационим и нерелационим

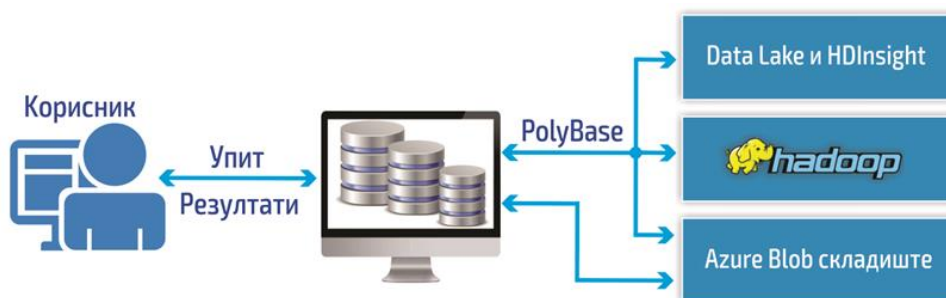
подацима унутар *Hadoop* оквира, који може да подржи оба, релациони и нерелациони податак паралелно (*DeWitt et al.*, 2013).

PolyBase је направљен у лабораторији *Jim Gray Systems Labs* коју је водио *David DeWitt*, професор Универзитета у Висконсину на Факултету рачунарских наука. *DeWitt* је познат по револуционарним истраживањима паралелних база података, интероперабилности независних база, објектно оријентисаног програмирања и *XML*-а.

PolyBase подржава више трећих страна (енгл. *Multiple Third Party*) *Hadoop* дистрибуције, укључујући *Linux*, *Cloudera* и *Hortonworks* платформу. *PolyBase* такође подржава и интеграцију пословне интелигенције са *Microsoft Office* апликацијама, *SQL* извештавање са сервера, и сервисе за напредне аналитике чиме се обезбеђује могућност задавања упита трећим странама *Hadoop* екосистема преко сродних алата за пословну интелигенцију (*Microsoft*, 2016):

- Извршава *T-SQL* упите над подацима смештеним у *Hadoop* екосистему;
- Извршава *T-SQL* упите над подацима смештеним на *Azure Blob* складишту;
- Извршава унос података у *Hadoop* екосистем, *Azure Blob* складиште или *Azure Data Lake* без додатних апликација и алата за засебне *ETL* процесе;
- Извршава извоз података из *Hadoop* екосистема, *Azure Blob* складишта или *Azure Data Lake*-а;
- Интегрише постојеће *Microsoft* алате за пословну интелигенцију, и примењује алате треће стране компатибилне са *SQL* сервером 2016.

PolyBase сервис пребацивања и измештања података (енгл. *PolyBase Data Movement Service – DMS*) ради са *HDInsight* и *HDFS* везом за паралелну и дистрибуирану обраду и процесирање комплексних нерелационих упита, побољшавајући перформансе и омогућавајући обраду података на месту њиховог постојања унутар складишта, без потребе за *ETL* процесима. Омогућава комбиновање релационих и нерелационих типова података, као и задавање упита над екстерним подацима који се налазе унутар *Hadoop* екосистема или *Azure Blob* складишта (приказано на слици 5) (*DeWitt et al.*, 2013; *Microsoft*, 2016).



Слика 5. Инфраструктура *PolyBase* технологије (адаптирано према: *Microsoft*, 2016)

Анализа структурираних и неструктурираних података је тешко применљива уколико не постоји начин за пренос података између различитих типова складишта. *PolyBase* решава повезивање екстерних података који су често смештени на *Cloud*-у, и унутар сервера. Упити над екстерним подацима могу да користе исту синтаксу као и упити над релационим базама података. Архитектура система може да подржи изградњу хибридних решења која испоручују увид у податке независно од тога где су складиштени (*DeWitt et al.*, 2013).

Инфраструктура је оптимизирана за паралелну обраду унутар више сервера чиме може да се обезбеди велика брзина рада. Сервери могу да подрже велики број процесорских језгара са векторским инструкцијама; могу да имају меморију изражену у ТВ и да омогуће високе улазно/излазне пропусне опсеге са локалним складиштем не нарушавајући њихову конзистентност унутар *Cloud*-а.

Azure складиште на *Cloud*-у усклађује исту *MPP* архитектуру дозвољавајући комбиновање скалирања снаге и еластичност облака. Еластичност као истакнута карактеристика система има способност за брзо ангажовање и ослобађање ресурса у зависности од захтева и обима задатог посла. Такав *Just in Time* систем може да оствари велике уштеде организацији (*Microsoft*, 2016; *Tejada*, 2017).

Big data инфраструктура и интеграција сервера са *Cloud* окружењем се остварује путем следећих сервиса и *big data* технологија (*Microsoft*, 2016):

- *Azure SQL Data Warehouse*;
- *Azure Data Lake*;
- *Azure HDInsight*;
- *Advanced Analytics & Power BI*.

У оквиру *Microsoft*-ове *big data* инфраструктуре постоје три сервиса (*Microsoft*, 2016):

- I. *Azure HDInsight* – Управља *Hadoop* кластерима и омогућава *Hadoop* сервисе на *Cloud*-у (*Chauhan et al.*, 2014). Решава питања у контексту истовременог приступа разноврсним подацима; скалирања ресурса са робусним алатима и сервисима користећи технологије за дељење добијених увида широм организације и географских локација (*Tejada*, 2017).
- II. *Azure Data Lake Analytics* – Дистрибуирани сервис за анализу података који је изграђен на *Apache YARN* платформи. Не захтева пројектовање захтевног хардвера. Упитима се трансформишу и извлаче вредносне информације из података при анализи, обради и визуализацији. Сервис за анализу података који може да изврши задати посао било које величине подешавањем и постављањем потребне конфигурације ресурса. Подржава активни директоријум на *Azure Cloud*-у и скалабилност дистрибуираног програмског језика *T-SQL* који

ефикасно анализира податке преко *SQL* сервера на *Azure* платформи, *Azure SQL* бази података и *Azure* складишту (*Ramakrishnan et al.*, 2017; *Klein*, 2017).

- III. *Azure Data Lake* – Омогућава инстанцу базе података у којој се могу сачувати подаци било које величине без извршавања промена унутар апликације. Пројектован је за процесирање података високих перформанси и аналитику *HDFS* апликација. Подаци могу бити генерисани у реалном времену са сензора, *IoT* уређаја или сајтова за електронску продају без фиксних ограничења величине (било да је кориснички налог или податак), дозвољавајући приступ највишег нивоа различитим подацима без њиховог померања или измештања аналитике. Предност задржавања података на месту где се налазе значи да иза успостављене везе није потребан додатни развој система (*Microsoft*, 2016; *Chauhan et al.*, 2014; *Tejada*, 2017; *Ramakrishnan et al.*, 2017; *Klein*, 2017).

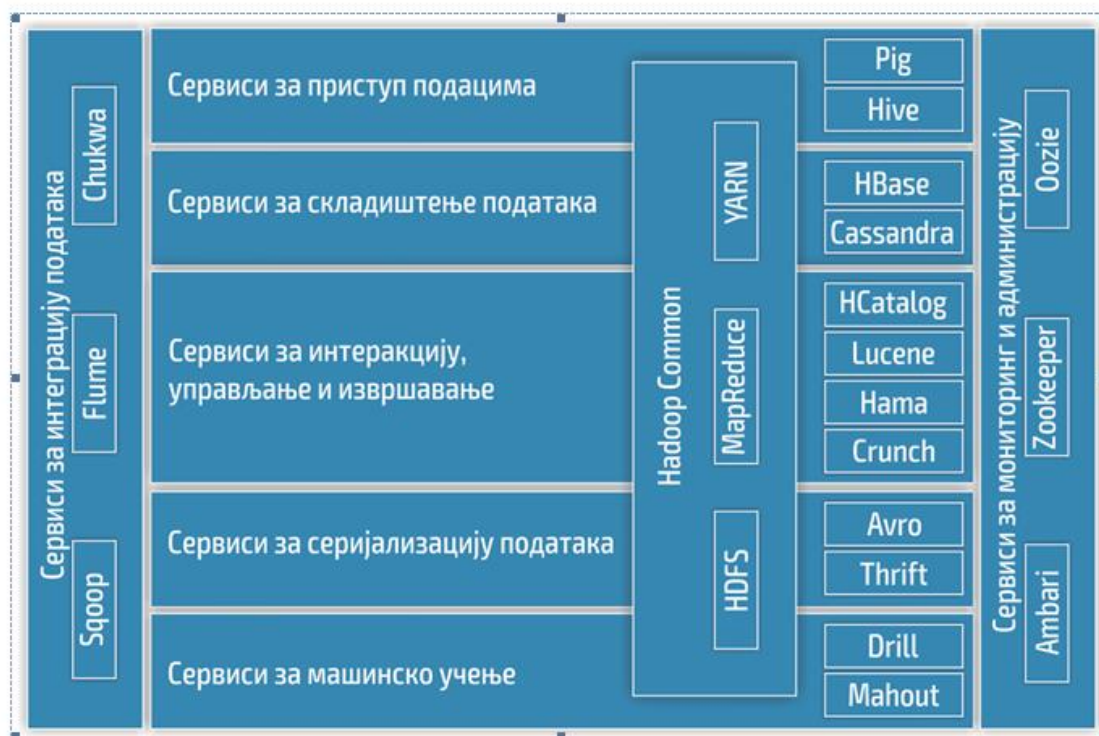
Применом *big data* технологија и реализацијом *big data* инфраструктуре, *Microsoft* је у сарадњи са *Intel* извршио значајан експеримент. Екосистем са складиштем од једног нода у кластеру величине 100TB, са четири процесора *Intel Xeon E7* на *SQL* серверу 2016, био је у стању да учита и припреми за обраду сложену *TPC-H* шему (енгл. *Composite Query-per-Hour Performance Metric*) за 1.6TB/h. Време израчунавања је било 5.3 секунди за извршавање сложеног упита у бази величине 100TB (*Microsoft*, 2016).

Иако и *MapReduce* и *PolyBase* технологије извршавају упите за паралелно процесирање података обезбеђујући скалабилну обраду релационих табела са неструктурираним подацима унутар *HDFS*-а, ова два система имају различите приступе. *Hadoop* користи *MapReduce* као основни процес за обављање паралелног процесирања (*Dean and Ghemawat*, 2008), док са друге стране *PolyBase* архитектура паралелно процесирање извршава применом *SQL* сервера 2016 (*Microsoft*, 2016). *MapReduce* примењује *HDFS* приликом *Shuffle and Sort* процеса, док *PolyBase* користи *Data Movement Service* за прераспodelу података преко нодова унутар кластера рачунара, и за унос и извоз података у/из *HDFS* складишта (*DeWitt et al.*, 2013).

3.1.4. **BIG DATA** СЕРВИСИ

Експоненцијалан раст укупне количине доступних података различитих типова и различитих извора је утицао на стварање *big data* парадигме. *Hadoop* је консолидована платформа која са пратећим развојним сервисима решава проблем анализе и обраде великих скупова података. Независне релационе базе података које постоје на тржишту и данас, остаће у употреби у годинама које долазе са задатком да решавају проблеме за које су намењени. *Hadoop* их допуњује и комплементаран је са постојећим системима обезбеђујући нови приступ обради и анализи где се структурирани подаци са сложеним, неструктурираним, могу комбиновати (*Olson*, 2010; *White*, 2012).

Hadoop може да садржи различите сервисе чијом се имплементацијом перформансе система побољшавају. Комбинујући сервисе са њиховим специфичним функцијама, реализује се ефикасан екосистем.



Слика 6. *Big data* сервиси – Технолошки оквир *Hadoop* екосистема

У центру се налази језгро система *Hadoop Common* – софтверски оквир за дистрибуирану обраду великих скупова података *YARN* (енгл. *Yet Another Resource Negotiator*). Пре добијања званичног имена, *YARN* се неформално звао *MapReduce2* или *NextGen MapReduce*. Међутим, уведен је нови приступ који је одвојио управљање ресурсима кластера и распоређивање послова од *MapReduce*-ове компоненте за обраду података, омогућавајући *Hadoop* екосистему да подржава различите типове обраде и шири спектар апликација. На пример, *Hadoop* кластери могу да покрећу интерактивне упите, аналитику и обраду података у реалном времену истовремено са *MapReduce* пословима (Vavilapalli et al., 2013).

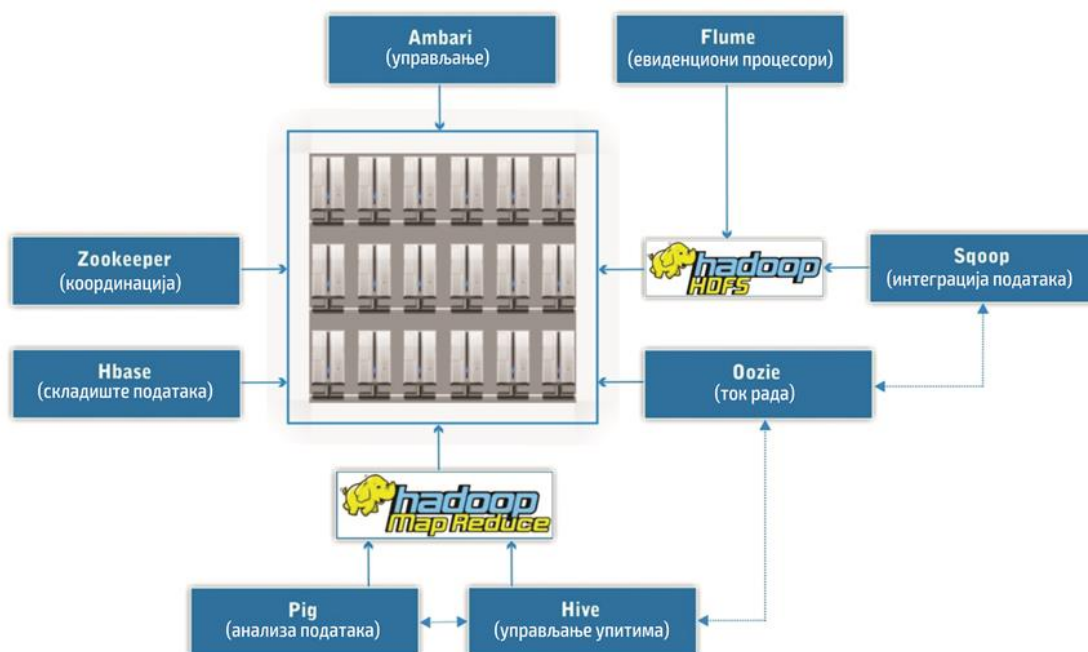
Технолошки оквир садржи следеће сервисе (Simović and Ćirović, 2016):

- Сервиси за приступ подацима. (1) *Pig* – сервис за анализирање великих скупова података који се састоји од програмског језика високог нивоа погодног за обављање анализе података заједно са инфраструктуром за вредновање програма. Истакнута особина *Pig* сервиса је структура подложна паралелизацији која омогућава управљање великим скуповима података. Скрипт-језик високог нивоа омогућава преузимање, читавање, трансформисање и филтрирање података, као што је *Pig Latin* – компајлер који производи секвенце *MapReduce* задатка које прослеђује кластеру; (2) *Hive* – управља упитима и великим

скуповима података смештених у дистрибуираном складишту. Складиште се односи на физичку локацију директоријума или датотека који су назначени *Hive* упитима. *Hive* обезбеђује механизам пројектовања структуре над подацима користећи језик сличан *SQL*-у – *HiveQL*. Истовремено, језик омогућава *MapReduce* програмирање када је неефикасно да се одређена логика изрази *HiveQL*-ом (White, 2012).

- Сервиси за складиштење података. *Hadoop* је пројектован за обраду и процесирање великих скупова података. Међутим, уколико је потребна претрага, преузимање и обрада специфичних података у реалном времену, то обезбеђују сервиси *HBase* и *Cassandra* (Dede et al., 2013). Циљ сервиса је складиштење веома великих табела, као што су примера ради, више милиона редова x милиони колона. Као што *Bigtable* усклађује складиштење дистрибуираним подацима добијених од *Google FS*-а (Chang et al., 2008), *Apache HBase* обезбеђује *Bigtable* попут способности *Google*-а. Термин *NoSQL* су омогућили *Hadoop* сервиси, дефинишући *Hadoop* екосистем флексибилним и скалабилним.
- Сервиси за интеракцију, управљање и извршавање. (1) *HCatalog* на менаџмент слоју *Hadoop* екосистема омогућава процесирање и обраду података помоћу сервиса *MapReduce*, *Pig* и *Hive*, као и лакше читање података са мреже. *HCatalog* апстрактно приказује преглед релационе базе података у *Hadoop* дистрибуираном систему датотека и осигурава чување података било ког формата (Pal, 2016); (2) *Lucene* – омогућава претраживање целог текста, проверу правописа, као и напредне анализе текста могућностима *API* алгоритама уграђених у токене; (3) *Hama* – осигурава израчунавање високих перформанси за извршавање алгоритама као што је *Matrix* и *Graph*. *Hama* обезбеђује *BSP – Bulk Synchronous Parallel* модел прослеђивања порука и колективне комуникације (Kajdanowicz et al., 2012); (4) *Crunch* – библиотеке које омогућавају оквир за писање, тестирање и покретање *MapReduce* процеса. Пружа пуну контролу над: мапирањем, редуковањем, сортирањем, комбиновањем, и помаже у извршавању агрегације података (White, 2012).
- Сервиси за серијализацију података (White, 2012) – метод који омогућава преузимање података из апликације и који обезбеђује процес превођења структуре података у формат који се може похранити на диску; накнадно разменити са другом апликацијом и реконструисати у њој разумљив формат (нпр. *JSON*, *XML* или други формат). (1) *Avro* – генерички *Hadoop* сервис за серијализацију и размену података; (2) *Thrift* – сервис за креирање флексибилних шема за рад са *Hadoop* подацима. Намењен је за вишејезичну компатибилност. Комбинује софтвер са изворним кодом и гради услугу која ради ефикасно са *C++*, *Java*, *Python*, *PHP*, *Ruby*, *Perl*, *C#*, *JavaScript* и другим језицима.

- Сервиси за машинско учење (*Landset et al., 2015*). *Drill* и *Mahout* – извршавају интерактивну анализу угњеждених података. Подржавају различите *NoSQL* базе података и системе датотека, укључујући *HBase*, *MongoDB*, *HDFS*, *MapR*, *Amazon S3* и друге. Један упит може да придружи податке из више независних база (*Zomaya and Sakr, 2017*). На пример, може да се придружи колекција профила корисника *MongoDB* са директоријумом серверских евиденција у *Hadoop* екосистему.
- Сервиси за интеграцију података. (1) *Sqoop* – сервис који обезбеђује интеграцију *Hadoop*-а са другим релационим системима. На пример, *MapReduce* резултат не мора да се смести унутар *HDFS*-а и да се обради *Pig* или *Hive* упитима, већ се може директно проследити у релационе системе. Такође, обезбеђује трансфер података из релационих система у *HDFS* (*Aravinth et al., 2015*); (2) *Flume* – евиденциони серверски процесори који у реалном времену обављају интеграцију са *Hadoop* екосистемом. Дистрибуиран, поуздан и доступан сервис за ефикасно прикупљање, агрегацију и трансфер великих евиденција сервера у *HDFS* (*Hoffman, 2013*); (3) *Chukwa* – сервис за прикупљање података и за праћење великих дистрибуираних система. Укључује флексибилан алат за приказивање, праћење и анализирање резултата ради избора сервиса за искоришћење прикупљених података (*Boulon et al., 2008; Zikopoulos and Eaton, 2011*).
- Сервиси за мониторинг и администрацију. (1) *Ambari* – омогућава инсталацију, управљање и мониторинг *Hadoop* сервиса у кластеру (*Wadkar and Siddalingaiah, 2014*); (2) *Zookeeper* – сервис за координацију дистрибуираних апликација. Синхронизује и централизује сервисе покренуте унутар кластера (*Hunt et al., 2010*); (3) *Oozie* – прати ток рада међусобно независних *Hadoop* сервиса (*Islam et al., 2012*).



Слика 7. Апстрактни приказ Hadoop сервиса

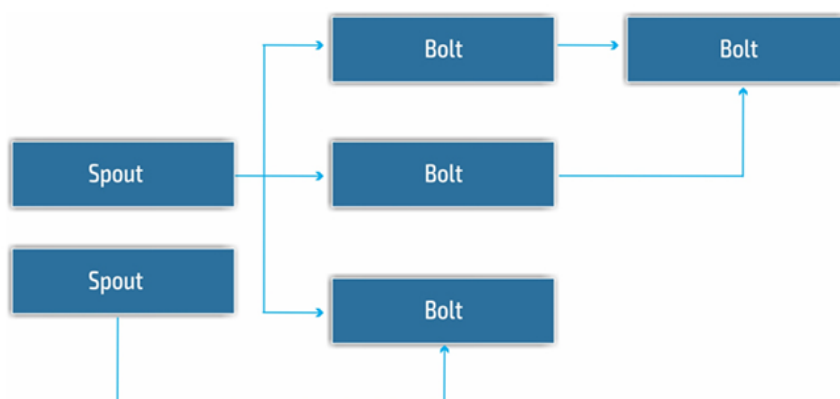
- Сервис *Storm*. Омогућава дистрибуирану обраду података у реалном времену (енгл. *Streaming*). *Storm* се може користити са било којим програмским језиком и омогућава аналитику у реалном времену, онлајн машинско учење и континуирано израчунавање. Има брз одзив и може да ради на преко милион торки (енгл. *Tuples*) које представљају модел података у *Storm* сервису. Реализован *Storm* кластер чине следећи чворови (*Leibiусky et al., 2012*):
 1. *Master Node – Nimbus*, који је одговоран за дистрибуирање података, додељивање задатака и мониторинг израчунавања;
 2. *ZooKeeper* чворови, који координирају рад кластера; и
 3. *Supervisor* чворови, који обављају комуникацију са *Master Node*-ом преко *ZooKeeper-a*; покрећу и гасе задатке додељене од стране *Master Node-a*.



Слика 8. Чворови сервиса Storm кластера

Рад се делегира на различите врсте компоненти које су одговорне за сваку специфичну обраду података у реалном времену. Улазни ток (енгл. *Stream*) сервиса *Storm* кластера обрађује прва компонента (енгл. *Spout*) која дефинише извор информација. *Spout* прослеђује податке компоненти *Bolt* која пристигле податке даље трансформише и

процесира. Да би се омогућило израчунавање у реалном времену дефинишу се топологије – графови израчунавања где сваки чвор у топологији садржи логику за обраду, и везе између чворова приказујући ток прослеђивања.



Слика 9. Топологија Storm сервиса у кластеру

Bolt може да апсорбује било који број улазних токова и да изврши њихову обраду. *Bolt*-ови могу да имају више функција: филтрирање торки; агрегацију токова и придруживање токова; комуникацију са базама података и др. *Spout* и *Bolt* компоненте су дефинисане у топологији као највиши ниво апстракције сервиса *Storm* кластера (Leibiusky et al., 2012). *Storm* сервис је брз и омогућава обраду више од милион записа у секунди на једном чвору у кластеру.

3.1.5. BIG DATA АНАЛИТИКА

Потенцијал *big data* технологија је значајан у дигиталним трансформацијама; води ка унапређењу комуникације између учесника и има стратешки значај за ојачавање кадровског потенцијала, модернизацију државе и убрзани привредни развој. Интеграција великих скупова података са новим технологијама и техникама визуализације може издвојити знање и створити могућност за решавање глобалних изазова које намећу нове информационе технологије (Simović, 2017). Способност система за примену нових технологија *big data* аналитике има потенцијал смањивања преоптерећења информација резултирајући новим научно-истраживачким решењима (Bertot and Choi, 2013).

Осим чињенице да се информационо-комуникационе технологије експоненцијално развијају, примена *big data* технологија у откривању нових знања је важан елемент за ефикасан приватни и државни сектор. У том редоследу, постоје сложени захтеви за пројектовање и имплементацију система машинског учења у модерним информационим системима. У *big data* аналитици, машинско учење као важно поље рачунарских наука има за циљ пројектовање алгоритама који омогућавају рачунарима да развијају понашање на основу постојећих података како би извршили прогнозе

будућих исхода и трендова (*Chen and Zhang, 2014*). Главна карактеристика машинског учења је развој модела који може открити обрасце података и аутоматски донети интелигентне одлуке (*Murphy, 2012*).

У практичној примени *big data* аналитике извршена је предикција догађаја коришћењем *Microsoft Azure Machine Learning* студија. Аналитичка решења, библиотеке и алгоритми који су спремни за употребу, обраду и анализу су обједињени *Cortana* галеријом (*Barga et al., 2015*) и могу се применити у *Microsoft Machine Learning* студију. *Azure ML* платформа обезбеђује алат за моделирање предиктивне аналитике и у потпуности пружа оперативан сервис у испоручивању прогностичких модела.

Узорак који је примењен у израчунавању је из доступне библиотеке података која је интегрисана у *Microsoft Azure Machine Learning* студију. Библиотеке података сачињавају спремне алгоритме за употребу који могу да се примене у изради модела за израчунавање и машинско учење. Узорак података се заснива, између осталог, на демографским карактеристикама, старосној доби, полу, образовању, занимању и друго, а у крајњој колони на годишњем приходу испитаника.

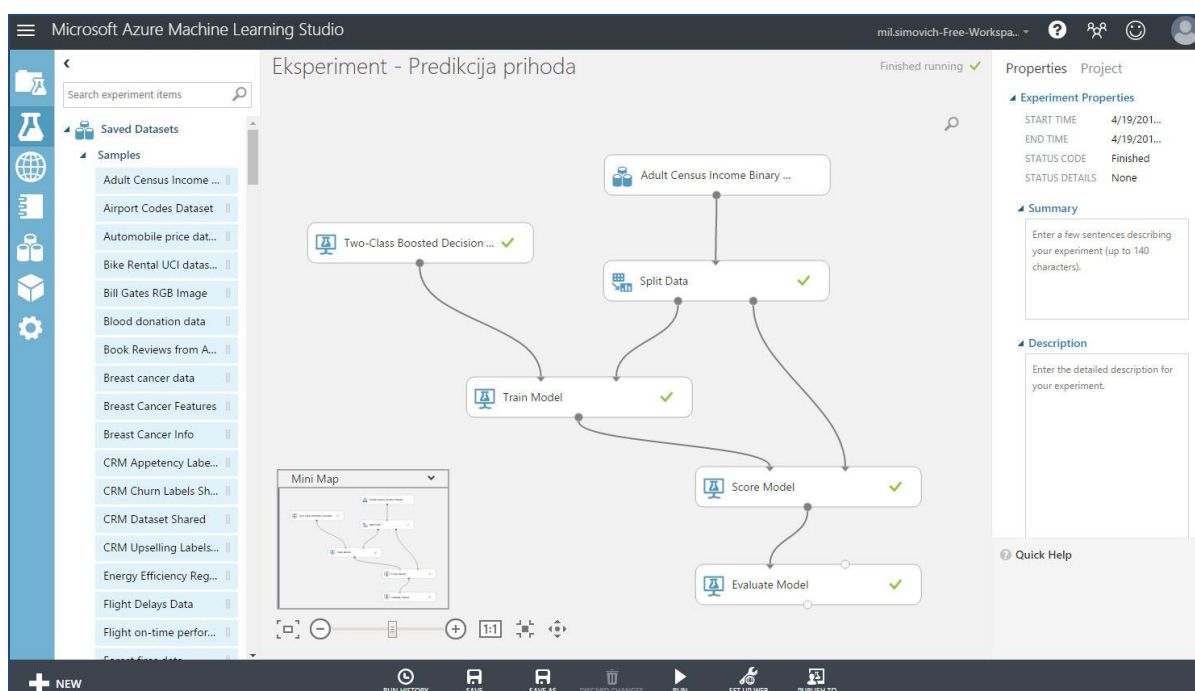
На слици 10 су приказане информације садржане у узорку. Коришћен скуп података има 32561 редова распоређених у 15 колона од којих је у 15-тој колони приказан годишњи приход (енгл. *Income*). Извршени експеримент треба да покаже тачност примењеног алгоритма и модела израчунавања машинског учења на основу података из доступне библиотеке *ML* студија.



Слика 10. Визуализација коришћеног скупа података

Експеримент је извршен на следећи начин:

- Извршена је подела скупа података на два дела. Први део од 70% је модел обуке, док је други део од 30% модел за тестирање;
- За интерконекцију модела обуке је изабран алгоритам *Two-Class Boosted Decision Tree* (Ту, 2005);
- Алгоритам обучава себе да би могао да предвиди приход заснован на информацијама примењеног скупа података;
- Модел обуке извршава израчунавање на примењеном скуп података;
- Креиран је модел резултата и модел евалуације;
- Урађене су међусобне везе модела обуке са заједничким подацима (алгоритам, подељени подаци и модел резултата);
- На крају је извршено коначно израчунавање и креиран је модел евалуације машинског учења (приказано на слици 11);
- Са успешно извршеним израчунавањем, модел машинског учења проширује скуп података и додаје две колоне моделу резултата (енгл. *Score Model*). То су: (1) постигнуте ознаке – приказано у колони 16 (енгл. *Scored Labels*) и (2) постигнуте вероватноће – приказано у колони 17 (енгл. *Scored Probabilities*).
- Проширени скуп података по извршеном израчунавању је приказан на слици 12.



Слика 11. Коначно израчунавање и креирање модела евалуације машинског учења

Eksperiment - Predikcija prihoda > Score Model > Scored dataset

rows: 9768 columns: 17

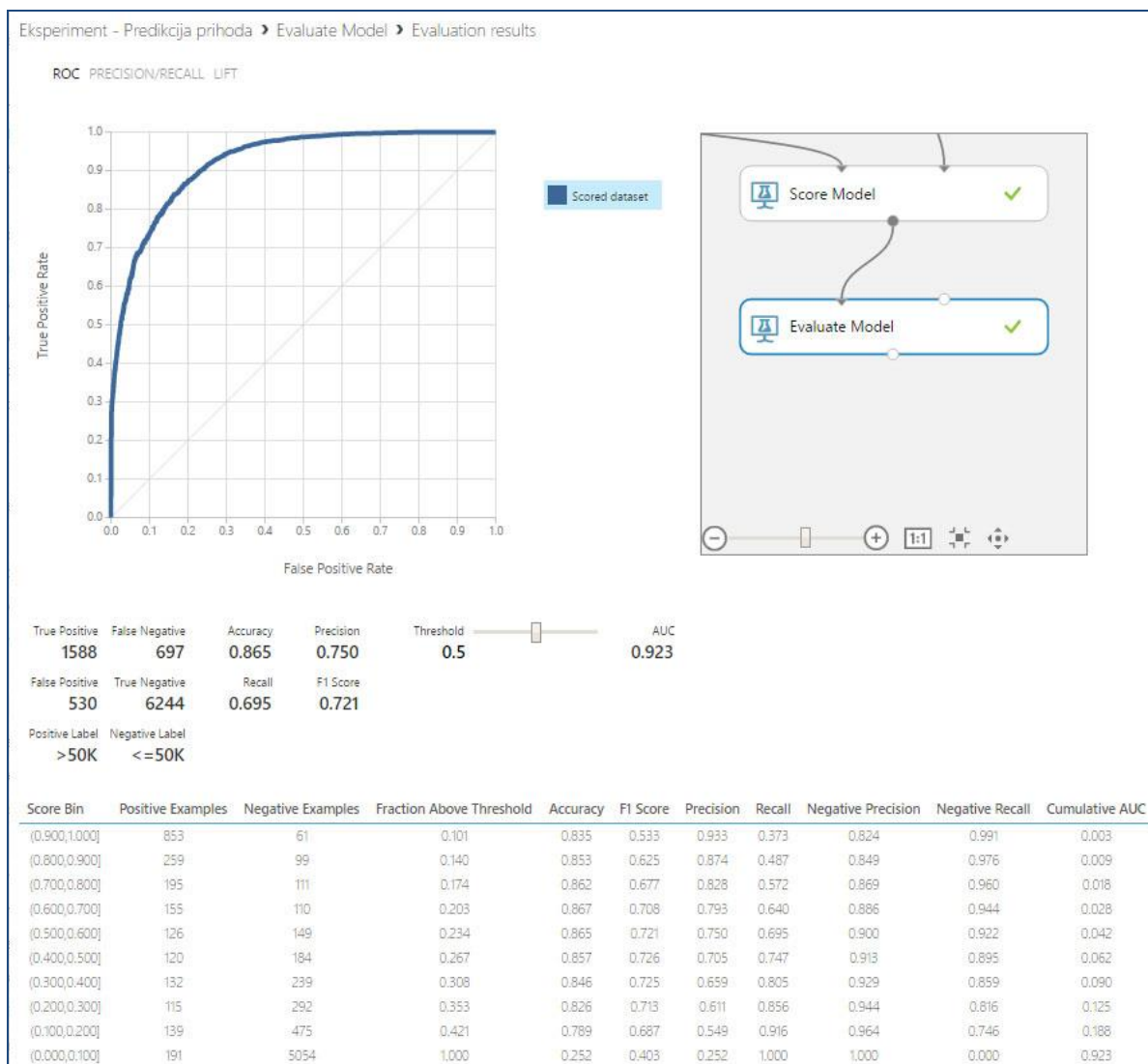
view as

age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	capital-gain	capital-loss	hours-per-week	native-country	income	Scored Labels	Scored Probabilities
37	Private	205339	HS-grad	9	Divorced	Craft-repair	Not-in-family	White	Male	0	0	49	United-States	<=50K	<=50K	0.038015
25	Private	394503	Some-college	10	Married-civ-spouse	Sales	Husband	White	Male	0	0	40	United-States	<=50K	<=50K	0.120344
56	Private	147989	Some-college	10	Married-spouse-absent	Transport-moving	Not-in-family	White	Male	0	0	40	United-States	<=50K	<=50K	0.053951
56	Private	169133	HS-grad	9	Married-civ-spouse	Other-service	Husband	White	Male	0	0	50	Yugoslavia	<=50K	<=50K	0.164289
57	Self-emp-inc	127728	Prof-school	15	Married-civ-spouse	Prof-specialty	Husband	White	Male	15024	0	60	United-States	>50K	>50K	0.999274
38	Private	23892	HS-grad	9	Married-civ-spouse	Machine-op-inspct	Wife	White	Female	0	0	40	United-States	<=50K	<=50K	0.328282
42	Self-emp-not-inc	436107	Assoc-acdm	12	Married-civ-spouse	Transport-moving	Husband	White	Male	0	0	50	United-States	<=50K	<=50K	0.276376
47	Private	102569	Bachelors	13	Married-civ-spouse	Adm-clerical	Husband	White	Male	0	0	40	United-States	>50K	>50K	0.64145
32	Self-emp-not-inc	298332	Bachelors	13	Married-civ-spouse	Prof-specialty	Husband	White	Male	0	0	45	United-States	>50K	>50K	0.577908
39	Private	435638	Some-college	10	Never-married	Machine-op-inspct	Not-in-family	White	Male	0	0	40	United-States	<=50K	<=50K	0.029381
40	Private	176716	Some-college	10	Married-civ-spouse	Sales	Husband	White	Male	0	0	40	United-States	>50K	>50K	0.507884
47	Private	357848	HS-grad	9	Married-civ-spouse	Craft-repair	Husband	White	Male	0	0	40	United-States	>50K	<=50K	0.442645

Слика 12. Визуализација проширеног скупа података

Анализа резултата проширеног скупа података и додатих колона је следећа:

- У првој додатој колони, постигнуте ознаке показују предвиђену вредност за сваког испитаника. Колона постигнуте ознаке се заснива на колони постигнуте вероватноће. Ако је вредност у колони постигнуте вероватноће већа од прага, она се предвиђа као позитивна (> 50К од годишњег прихода). У супротном се предвиђа као негативна (<= 50К од годишњег прихода);
- У другој додатој колони, постигнуте вероватноће, вредности показују вероватноћу да испитаник припада позитивној класи. Што значи да ако је први број у колони, примера ради 0.038015, постоји 0.038015 вероватноћа да први испитаник припада позитивној класи на *ROC* криви.



Слика 13. Резултати евалуације експеримента big data аналитике

У анализи и евалуацији експеримента изведеног коришћењем *ML* студија приказана је *ROC* крива (енгл. *Receiver Operating Characteristic Curve*). *ROC* крива илуструје способност система бинарног класификатора као праг дискриминације (приказано на слици 13) и представља однос тачних позитивних резултата као вероватноћу детекције (енгл. *True Positive – TP*) и нетачних позитивних резултата као лажни аларм (енгл. *False Positive – FP*). У изведеном експерименту *big data* аналитике, тачне позитивне вредности и вероватноћа детекције, показали су квалитетне резултате јер је сензитивност у *ROC* простору веома висока.

AUC крива (енгл. *Area Under the Curve*) је са друге стране приказала нормализовање резултата, површину испод криве, и представља вероватноћу да ће пројектовани модел насумично рангирати изабране позитивне инстанце које су веће од насумично одабране негативне инстанце под претпоставком – позитивни опсег испред негативног, где је у извршеном експерименту вредност од 0.923 – скоро идеалан резултат.

Такође, у изведеном експерименту *big data* аналитике даља анализа елемената добијених резултата обухвата:

- Тачност (енгл. *Accuracy*), која мери квалитет модела класификације као проценат позитивних резултата у укупним случајевима, је 0,865, што је близу идеалног;
- Прецизност (енгл. *Precision*), која је удео укупно позитивних резултата, је 0.750, и веома је висока;
- Оповив (енгл. *Recall*), који је део свих позитивних резултата, и износи 0.695;
- Ф-запис (енгл. *F-score*), који се израчунава као пондерисани просек прецизности и оповива, што је у резултату евалуације 0,721 (идеална вредност Ф-записа = 1).

Постоје могућа ограничења која се односе на анализу резултата изведеног експеримента *big data* аналитике:

1. Скуп података који је коришћен за примену предиктивног аналитичког решења је јавно доступан, интегрисан је у представљени оквир и даје високе резултате. Коришћење физички прикупљених података би резултовало знатно нижим вредностима;
2. изабрани алгоритам који је примењен у израчунавању неће моћи да обради веће скупове података јер метод омогућава учење и обуку тако што друго стабло исправља грешке првог, треће стабло исправља грешке првог и другог и тако даље.

У том редоследу, ако је количина и разноврсност скупа података који је спреман за обраду већа, постоје веће могућности за проналазак скривених увида у обрасце (*Simović, 2017*).

3.2. СИСТЕМИ ПРЕПОРУКА

Системи препорука дефинишу скуп садржаја који за активног корисника могу бити потенцијално корисни и који се налазе у подручју његовог интереса. У односу на начин генерисања, систем кориснику може понудити једну или више препорука. Да би био у могућности да одреди релевантну препоруку, систем треба да уважава више критеријума за представљање корисности садржаја базираних на проширеном скупу података о кориснику и ставкама система који се препоручују (*Adomavicius and Tuzhilin, 2005; Simović, 2014; Bernardes et al., 2014; Davidović et al., 2018*).

Обзиром на технике и алгоритме који се примењују, основни системи препорука се могу поделити на (*Adomavicius and Tuzhilin, 2005; Lu et al., 2015; Ricci et al., 2011*):

- системе препорука базиране на садржају (енгл. *Content Based Recommendations*);

- системе колаборативног филтрирања (енгл. *Collaborative Filtering*);
- хибридне системе препорука (енгл. *Hybrid Recommender Systems*).

3.2.1. СИСТЕМИ БАЗИРАНИ НА САДРЖАЈУ

Код система препорука базираних на садржају, приступ се заснива на претпоставци да ће корисност производа бити сличних карактеристика онима које је корисник бирао у прошлости. На основу претходно оцењеног или одабраног садржаја, систем настоји да предвиди преференције корисника ослањајући се на аутоматизована правила која се користе при обради. Системи базирани на садржају омогућавају да се одреде правила која су често заснована на демографским или другим карактеристикама корисника (Pazzani, 1999; Lu et al., 2015).

Приликом генерисања препоруке, преференције корисника се утврђују на бази карактеристика садржаја које је одабрао или оценио. Ти подаци се упоређују са карактеристикама осталих ставки и издвајају садржаји који се по карактеристикама у највећој мери подударају са преференцијама корисника. Све акције корисника које систем препознаје су подељене у две категорије: (1) експлицитне и (2) имплицитне (приказано на слици 14). Експлицитне су интеракције корисника и система у вези корисности садржаја путем оцењивања, коментарисања или гласања; док су имплицитне, интеракције корисника са системом путем анализе акција и корисних радњи, *clickstream* анализе, куповине и праћења (Schafar et al., 1999; Lops et al., 2011).

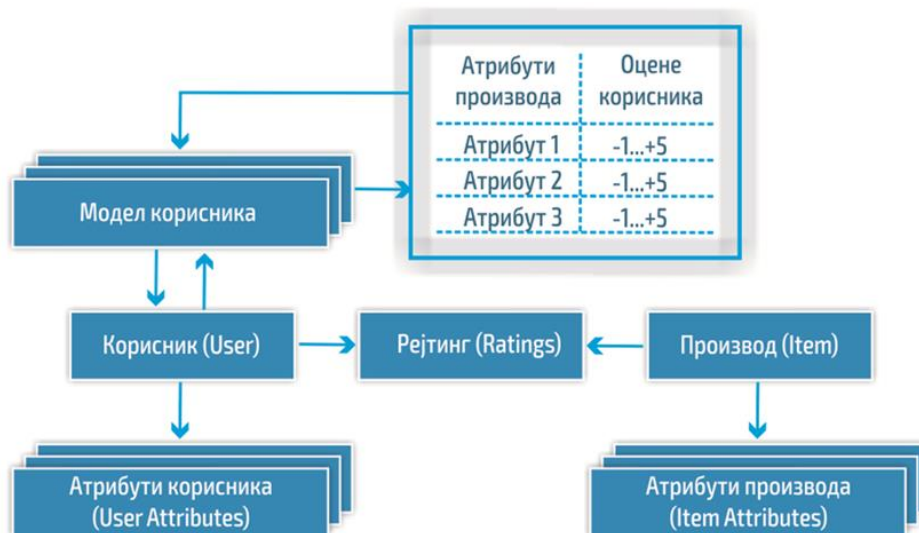


Слика 14. Основни модел преференција корисника

Оцена представља меру којом се исказује или предвиђа колико је за циљаног корисника одређени садржај релевантан. Да би то било могуће, сваки се корисник описује моделом који може да садржи, како основне податке о самом кориснику и податке о његовим циљевима, тако и његове преференције уопште. Такође, садржаји се описују скупом карактеристика и структурираним атрибутима који представљају модел производа (Burke, 2002; Koren and Bell, 2015).

Први корак у процесу генерисања препоруке је изградња модела корисника. Потребни подаци се притом прикупљају од корисника (нпр. путем упитника) или се аутоматски одређују на бази интеракције корисника са системом. У следећој фази, подаци се из модела упоређују са карактеристикама осталих производа и подацима о осталим корисницима система како би се одредила релевантна препорука (*Burke, 2002; Koren and Bell, 2015; Ricci et al., 2011*).

Код система препорука базираних на садржају, корисник оцењује производе чиме се креира модел корисника у односу на атрибуте оцењених производа и висине датих оцена (приказано на слици 15).



Слика 15. Систем препоруке базиран на садржају

Модел корисника је приказан кроз табелу вектора речи (енгл. *Keyword Vector*). Сваки производ има своју позицију у векторском простору. Позицију дефинише вектор који се рефлектује кроз опис, кључне речи и атрибуте, на позицију у том простору. Корисник се описује моделом преференција који такође припада простору вектора. Поклапање између преференција корисника и посматраног производа је мера близине тих вектора (*Salton et al., 1975; Simović, 2014*).

У односу на начин приказа препоручених садржаја, кориснику се може понудити једна или више препорука (нпр. препоруке поређане према висини оцена). Неки од постојећих система препорука користи један критеријум за представљање корисности производа док се у новијим истраживањима указује на важност уважавања више критеријума како би се одредиле релевантније препоруке. Притом се препорука генерише на темељу проширеног скупа података о кориснику и производима који се препоручују (*Lops et al., 2011*).

Једно од ограничења система препорука базираних на садржају је ограничена могућност аутоматске анализе садржаја. Успешност рада система препорука базираних

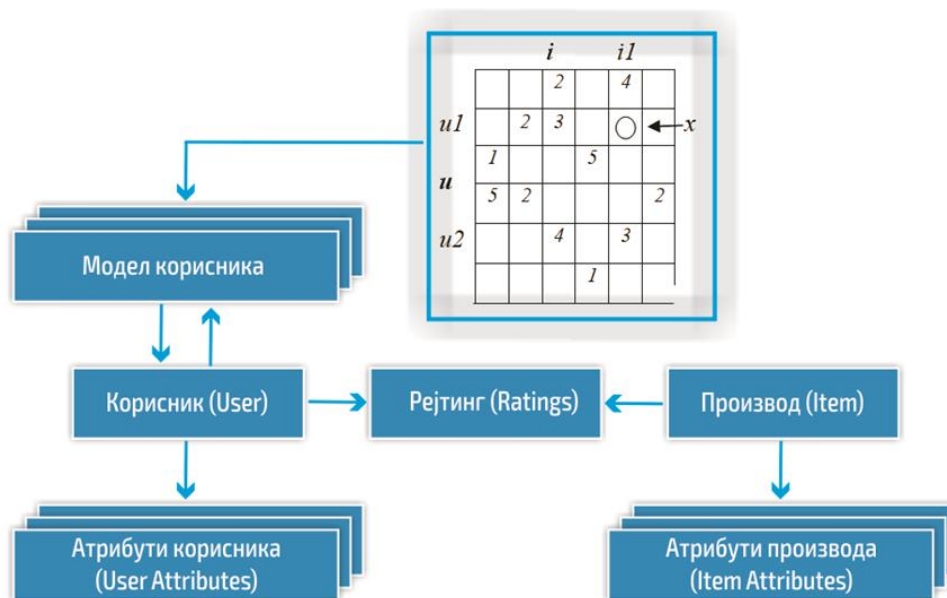
на садржају зависи од доступних података којима се описују производи, тако да је из тог разлога потребно омогућити аутоматско описивање производа или њихове карактеристике придружити ручно. Аутоматско додељивање карактеристика производа је у неким случајевима сложено, нпр. код описа мултимедијалног садржаја, док је ручни опис временски захтеван (*Schafer et al.*, 1999; *Burke*, 2002).

Ограничење система препорука базираних на садржају је и хладан старт (енгл. *Cold Start*) (*Schein et al.*, 2002; *Guo et al.*, 2014). Због начина рада система, потребно је да корисник оцени довољно садржаја како би систем имао релевантне податке и могао да генерише одговарајуће препоруке. Такође, недостатак система препорука базираних на садржају је тај што препоручени садржај често није разноврстан. Због начина одређивања корисности производа, где се упоређују карактеристике производа са профилем корисника и препоруке садржаја за које је та вредност максимална или веома висока, догађа се да се кориснику никада не препоручују производи који се разликују по садржају од онога што је корисник прегледао раније (енгл. *Sparsity*) (*Adomavicius, and Tuzhilin*, 2005; *Ricci et al.*, 2011; *Guo et al.*, 2014).

3.2.2. СИСТЕМИ КОЛАБОРАТИВНОГ ФИЛТРИРАЊА

Системи колаборативног филтрирања настоје да предвиде корисност производа на основу вредновања тог производа од стране осталих корисника система. Код система колаборативног филтрирања, оцене које су корисници доделили производима користе се као апроксимативна презентација њихових интереса и потреба (*Simović*, 2014).

За разлику од препорука на основу садржаја, модел корисника се креира на основу доступних података о производима које је оценио. Оцене се даље упоређују са оценама које су додељене од стране осталих корисника система ради одређивања најближег скупа производа од интереса. Посматраном кориснику ће међу производима које још није оценио бити препоручени они које су високо оценили њему слични корисници. Приступи колаборативног филтрирања се разликују управо по начину на који се одређује сличност између корисника (*Herlocker et al.*, 2004).



Слика 16. Систем колаборативног филтрирања

Као што је приказано на слици 16, за корисника $u1$ систем треба да да препоруку тако што ће проценити тражени производ $i1$ из листе непотпуних података (енгл. *Missing Value*) (Kagie et al., 2011).

Системи колаборативног филтрирања се класификују у два основна приступа (Sarwar et al., 2001; Herlocker et al., 2004):

- колаборативно филтрирање према корисницима (енгл. *User Based*);
- колаборативно филтрирање према производима (енгл. *Item Based*).

Код колаборативног филтрирања према корисницима, вредновање производа се упоређује са вредновањем осталих корисника система како би се одредила група сличних корисника (енгл. *Neighborhood*) (Sarwar et al., 2001). Препоручују се производи које су други корисници оценили највишим оценама. На тај начин се посматраном кориснику препоручују садржаји које преферирају њему слични корисници.

Са друге стране, код колаборативног филтрирања према производима, за садржаје које је посматрани корисник прегледао, купио или оценио, проналазе се и препоручују слични садржаји. Приликом одређивања сличности садржаја упоређује се вредновање посматраног корисника са вредновањем осталих корисника система (Herlocker et al., 2004). Циљаном кориснику се препоручују садржаји који су остали корисници приближно вредновали.

У оба наведена случаја, скуп садржаја за препоруку је ограничен на оне производе које су оценили остали корисници. Системи колаборативног филтрирања решавају неке проблеме уочене код система препорука базиране на садржају. Препоруке су независне

у односу на садржај јер се приликом генерисања препорука користе оцене осталих корисника, а не карактеристике садржаја који се препоручују (*Adomavicius, and Tuzhilin, 2005; Ricci et al., 2011*). Такав начин рада омогућава да се кориснику препоруче производи који се садржајно разликују од онога што је корисник прегледао раније.

Међутим, системи препорука код којих се примењују само технике колаборативног филтрирања такође имају одређене недостатке. Један од њих је проблем малог броја оцена (енгл. *Sparsity*) који је нарочито присутан код система са великим бројем производа које је могуће препоручити (*Guo et al., 2014*). Како успешност система колаборативног филтрирања зависи од упоређивања производа који су вредновани од стране корисника, проблем се јавља када је доступан мали број оцена (*Koren and Bell, 2015*). Производи који су вредновани од стране малог броја корисника готово никада неће бити препоручени, независно од висине датих оцена. Такође, ако се скуп садржаја неког производа често мења, раније додељене оцене неће користити новим корисницима јер ти производи неће бити обухваћени алгоритмом за генерисање препоруке.

И код колаборативног филтрирања се јавља проблем хладног старта за нове кориснике. Уколико на почетку интеракције са системом нема оцењених садржаја од стране корисника, није могуће генерисати одговарајуће препоруке. Исти се проблем јавља и у случају новог садржаја у систему. Све док нови садржај не оцени довољан број корисника, систем га неће препоручивати. Тај проблем је посебно изражен код система где постоји сталан унос новог садржаја (*Schein et al., 2002*). Како се код колаборативног филтрирања знање система о производима заснива на томе шта корисници преферирају, препоруке зависе у потпуности од тога које садржаје корисници оцењују, што резултира препоручивањем само оцењених и вреднованих производа од стране корисника система (*Ricci et al., 2011*).

Корисници чија се интересовања разликују у односу на остале учеснике, и за које је тешко наћи сличности, не могу од система препорука колаборативног филтрирања очекивати добре резултате. Колаборативно филтрирање се показало успешно за оне кориснике који се могу сврстати у групу најближих корисника, док за остале кориснике није могуће понудити одговарајуће препоруке. Уз наведено, постоји и проблем разноврсности садржаја (енгл. *Diversity*) (*Shani and Gunawardana, 2011*). Како се код колаборативног филтрирања по производима, знање система заснива на томе шта корисници преферирају, препоруке зависе у потпуности од тога које садржаје корисници оцењују, што резултира препоручивањем само најпопуларнијих садржаја (*Burke et al., 2015*).

Један од недостатака колаборативног филтрирања је флексибилност (*Beliakov et al., 2011*). Већина система подржава генерисање препоруке зависно од вредновања према једном критеријуму, али не и препоруке на основу више критеријума. Такође, у већини случајева производи се препоручују појединцима али не и групама (*McCarthy et al., 2006; Masthoff, 2011*). Проблем је и скалабилност (*Shani and Gunawardana, 2011*).

Приликом коришћења методе попут алгоритма за проналажење најближих производа, систем се не прилагођава повећању броја корисника.

Колаборативно филтрирање даје три предности филтрирања информација приликом генерисања препоруке у односу на системе базирани на садржају: (а) филтрирање производа чији садржај није анализиран аутоматизованим процесима; (б) филтрирање производа на основу квалитета; и (в) могућност случајног открића корисног производа у генерисаној препоруци која може бити препорука од интереса за корисника а која у садржају није описана очекиваним концептом (енгл. *Serendipitous*) (*Herlocker et al.*, 2004).

Могућност колаборативног филтрирања у релевантнијем препоручивању је велика. Међутим, у достизању пуног потенцијала потребно да се комбинује са технологијом система препоруке базираном на садржају. Колаборативно филтрирање само по себи остварује добро предвиђање оних производа који су за корисника од интереса, али не и за проналажење специфичног контекста у информационом систему. Проблем аутоматизованог колаборативног филтрирања информација је предвиђање колико ће кориснику одговарати онај производ са којим није упознат, а који има скуп преференција осталих корисника система (*Ricci et al.*, 2011).

Код колаборативног филтрирања, за предвиђање систем бележи укупно вредновање посматраног производа, укључујући и модел корисника. Активан корисник омогућава систему да кроз изабран производ добије листу препоручених у простору који може бити формулисан као матрица корисника и производа, где свака ћелија представља вредновање корисника за специфичан производ (*Herlocker et al.*, 1999) (поједностављен приказ матрице је приказан на слици 17).

Матрица	<i>i1</i>	<i>i2</i>	<i>i3</i>	<i>i4</i>
<i>u1</i>	5	2	5	4
<i>u2</i>	2	5		3
<i>u3</i>	2	2	4	2
<i>u4</i>	5	1	5	?

Слика 17. Матрица корисника и производа

Колаборативно филтрирање се репрезентује кроз предвиђање непостојећих вредности у матрици. Нпр, систем треба да омогући кориснику *u4* предикцију за производ *i4*, односно треба да израчуна вредност празне ћелије. Код колаборативног филтрирања, матрица може бити изузетно ретка (енгл. *Sparsity*) обзиром да је сваки корисник оценио

мали проценат од укупног броја производа (Sarwar et al., 2001; Guo et al., 2014). У колаборативном филтрирању, метод упоређивања сличности корисника (енгл. *Neighborhood*) (Herlocker et al., 1999) има за циљ упоређивање сличности корисника у односу на вредновање производа приликом предвиђања:

$$P_{a,i} = \frac{\sum_{u=1}^n r_{u,i}}{n} \quad (1)$$

P је предвиђање за корисника a и производ i , које се може израчунати сумирањем вредновања скупа n корисника за исти посматрани производ i . Сви посматрани корисници скупа су оценили производ i за посматраног корисника a , коме је предикција P намењена (приказано у мат. формули 1). Међутим, ако корисник a није оценио производ, није укључен у сумирање вредновања скупа n .

Персонализацију ће систем обавити боље (мат. формула 2), када предикцију P за корисника a и производ i , у скупу n корисника и свих производа i , не посматра једнако, већ суми n корисника и рејтингу корисника u за посматрани производ i додаје тежину персонализације која се рефлектује кроз договор (енгл. *Agreement*) корисника a и корисника u ($\omega_{a,u}$) у односу на суму тих корисника и тежине њихове персонализације.

$$P_{a,i} = \frac{\sum_{u=1}^n r_{u,i} * \omega_{a,u}}{\sum_{u=1}^n \omega_{a,u}} \quad (2)$$

Интуитивно, ако је тежина персонализације изражена кроз 0 или 1 , односно зависна од близине корисника (енгл. *Neighbourhood*), израз ($\omega_{a,u}$) ће вратити резултате који ће приказати колико је корисника укључено у предикцију израчунавања просека (енгл. *Average*) корисника сличних преференција у односу на активног корисника.

Међутим, тежина персонализације не мора бити вреднована кроз 0 или 1 , већ може бити од 0 до 1 , где је вредност 1 најбоље поклапање, док остале вредности које су ближе 0 мање учествују (или уопште не учествују) у генерисању просека активног корисника.

Начин да се приликом израчунавања побољша препорука (мат. формула 3) је да систем препозна да вредновање није на истој скали:

$$P_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_u)}{n} \quad (3)$$

Предвиђање за посматраног корисника a за производ i , чак иако нема додељену тежину персонализације биће сума корисника система од u до n ; вредновање које додељује

корисник u за производ i , минус просечна вредност корисника u , чему је додата девијација (енгл. *Deviation*) просека корисника a .

Када се додају тежине (енгл. *Weights*), путем Пирсонове корелације за поређење сличности укључивањем свих корисника, израчунавање финалне предикције представља тежина просека девијације:

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_u) * \omega_{a,u}}{\sum_{u=1}^n \omega_{a,u}} \quad (4)$$

Када су вредности персонализације додате, предикцију за корисника a за посматрани производ i , систем израчунава тако што упоређује како су сви корисници система вредновали посматрани производ корисника a у односу на производе које су сви корисници оценили, и колико их систем вреднује кроз тежину персонализације посматраног корисника у односу на остале кориснике система.

Комбинујући те елементе у односу на тежину персонализације добијају се обједињене вредности тежине персонализације свих корисника система и корисности производа од интереса. Добијени резултати се упоређују са просеком посматраног корисника након чега систем генерише препоруку.

Међутим, приказана формула није математички прецизна јер прелази дозвољене границе (енгл. *Out of Range*). Нпр, уколико систем рангира производе од 1 до 5*; а просек посматраног корисника је 4,5* и тражи препоруку најпопуларнијег производа који је оцењен високо (нпр. са 4 или 5*) од стране корисника чији је просек 2 или 3*; предикција система за посматрани производ ће бити ≈ 7 од 5*.

Финално предвиђање (мат. формула 5) се извршава путем израчунавања коефицијента Пирсонове корелације узимајући у обзир тежину просека девијација сличних корисника (енгл. *Pearson Correlation Coefficient*):

$$\omega_{a,u} = \frac{\sum_{i=1}^m (r_{a,i} - \bar{r}_a) * (r_{u,i} - \bar{r}_u)}{\sigma_a * \sigma_u} \quad (5)$$

Израз потврђује претпоставку да систем треба да да веће вредности производима који су високо оцењени, нарочито ако се слажу са екстремним производима, односно оним производима који су оцењени веома ниским оценама (*Herlocker et al.*, 1999).

3.2.3. ХИБРИДНИ СИСТЕМИ

Хибридни системи комбинују алгоритме засноване на садржају и колаборативном филтрирању и настоје да превладају потешкоће и ограничења појединачних приступа. Хибридни системи узимају у обзир карактеристике садржаја као и вредновање садржаја од стране корисника. Приступи имплементацији зависе од начина комбиновања различитих алгоритама за генерисање препорука. Неке од могућности су да се на основу одређеног критеријума примењује више техника у неколико фаза. Да се применом једне технике добије модел који служи као улаз за следећу технику, или да се заједно презентују препоруке добијене на основу различитих техника (*Burke, 2002; Ricci et al., 2011*). Методе комбиновања техника у хибридним системима су (*Burke, 2002*):

- Хибридно генерисање. Генерисане препоруке применом различитих система су представљене активном кориснику једна до друге кроз комбиноване листе. Листе су одвојене и не генеришу се у једну.
- Хибридно преклапање. Представља хибридно генерисање препоруке на основу тренутног прегледа производа од стране активног корисника на основу унапред дефинисаног критеријума о избору листе производа за препоручивање.
- Додељена тежина персонализације посматраног модела корисника. Представља комбинацију више компоненти за генерисање препоруке која се нумерички комбинује. За оцењене производе, резултати се линеарно упоређују са моделима корисника система.
- Хибридно побољшање. Представља комбинацију више техника за генерисање препоруке. Једном техником се израчунава скуп карактеристика производа у систему и ажурирају се модели, док се са другом техником израчунавају сличности између корисника система и упоређују њихове оцене.
- Комбинација више хибридних система. Представља генерисање препоруке из различитих извора које се затим као улазне вредности комбинују са алгоритмом који генерише јединствену персонализовану препоруку.

Дефинисање персонализације у контексту система препоруке је процес којим се настоји што је могуће боље прилагодити одређену препоруку активном кориснику. Резултат рада хибридног система су препоруке за чије генерисање се анализирају како подаци о корисницима тако и садржаји који се препоручују.

3.2.4. ПЕРСОНАЛИЗАЦИЈА СИСТЕМА ПРЕПОРУКА

Системи препорука дефинишу скуп садржаја персонализујући искуство корисника (нпр. производе за куповину, филмове за гледање, текстове за читање, људе за праћење, и др.). Персонализација у системима препоруке представља способност прилагођавања и препоручивања садржаја у складу са корисничким преференцијама. Персонализација се може посматрати као предвиђање проблема, где систем треба да предвиди ниво интересовања или корисности одређене категорије садржаја, веб странице или других елемената и да их рангира према предвиђеним вредностима (*Schafer et al.*, 1999). Способност персонализације у генерисању препоруке, и његова могућност да прилагоди садржај и да препоручи адекватне производе, подразумева да систем може да закључи шта корисник тражи на основу претходних итерација као и тренутне интеракције са активним корисником. Основни задатак персонализације је предвиђање нивоа интересовања корисника, корисност одређеног садржаја, и коначно рангирање у складу са највишим интересним вредностима производа или услуге за активног корисника (*Schafer et al.*, 2001; *Lu et al.*, 2015).

У основи се персонализација система може приказати као мапирање корисника и производа у скупу интересних вредности. Персонализација као функција предвиђања произилази из чињенице да мапирање није дефинисано у целом домену доступности садржаја, те самим тим захтева од система процену вредности интересовања активног корисника за поједини садржај (*Schafer et al.*, 1999; *Schafer et al.*, 2001; *Balabanović and Shoham*, 1997).

Системи препорука генеришу садржај различитог нивоа персонализације. Ниво персонализације обухвата неколико фактора укључујући и тачност и корисност препоруке. Оба мерења, и тачност и корисност, која укључују факторе као што је случајан узорак могуће вредности производа, су важна. Систем који препоручује само најпродаваније производе има мању прецизност од система препоруке који може да пронађе и препоручи више производа од интереса одређеном кориснику. Сличност се не може увидети код система препоруке који препоручује удаљене, различите производе, што је неретко далеко од интересне вредности кориснику.

Док је персонализација континуум више димензија, корисно је идентификовати заједничке нивое. Када системи препоруке обезбеде идентичне препоруке сваком активном кориснику, систем је класификован као неперсонализован. Специфичне и тачно одређене препоруке могу се заснивати на ручној селекцији, статистичкој сумаризацији или другим техникама. Чести примери система препорука у пракси су неперсонализовани. Најпродаванији производи, избори администратора или модератора, просечан рејтинг и нефилтрирани коментари купаца, сви представљају исте или насумице изабране препоруке сваком кориснику система.

Системи препоруке који користе текуће улазне параметре корисника како би прилагодили препоруку од интереса, представљају краткорочну (енгл. *Ephemeral*)

персонализацију. Међутим, ти системи представљају корак више у односу на неперсонализоване системе јер обезбеђују препоруке које одговарају навигацији и селекцији купца кроз систем. Систем препоруке са вишим нивоом краткорочне персонализације би био онај који користи целу претраживачку сесију или садржај корпе за куповину активног купца како би препоручио адекватне производе.

Дубоко персонализовани (енгл. *Deep Personalization*) (Schafer et al., 2001) системи препорука генеришу другачије листе за различите купце чак и када они гледају исте производе у оквиру једног система. Ови системи користе корелацију између корисника и препоруке засноване на особинама користећи предности постојећих преференција и корелацију између производа. Дубока персонализација има за циљ да развије дугорочну везу са купцима што резултира кредибилитетом. Заснована је на историји обављених куповина кроз преференције, навигацију и других параметара. Коришћењем колаборативних могућности, пореди се историја одређеног купца са историјом осталих купаца у систему и генеришу се персонализоване сугестије и предикције. Дубока персонализација гради систем управљања односима са клијентима током времена, остављајући историју обављених куповина развијеном, притом обезбеђујући побољшање будућих препорука. За разлику од система који захтевају ручно ажурирање, дубоко персонализовани систем са особинама учења у току рада може да користи корелацију између корисника система и да анализом атрибута идентификује производе од интереса (Schafer et al., 2001).

Са пријавом корисника на систем (приказано на слици 18), улазне информације се обрађују ради ажурирања модела и генерисања персонализованих препорука. Апликације које не користе улазне информације о циљаном кориснику могу да произведу само неперсонализоване препоруке. Комбиновање и додавање више типова улазних параметара омогућава систему да прилагоди препоруке на основу тренутне активности и дугорочне преференције корисника.



Слика 18. Модел корисника у персонализованом систему

Персонализовани системи не замењују корисника на вебу нити обавезно доносе одлуке, већ је њихова улога да прикупе, организују и представе кориснику скуп релевантних информација и омогуће интересни садржај.

3.2.5. ИНТЕЛИГЕНТНИ АГЕНТИ У СИСТЕМИМА ПРЕПОРУКА

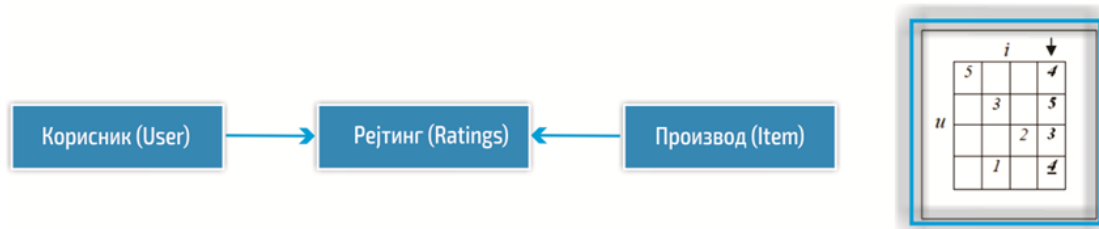
Интелигентни агенти су софтверска решења пројектована да самостално или на основу захтева корисника обављају одређене акције на вебу. Агенти имају могућност филтрирања информација, обраде и повезивања доступних података.

Интелигентни агенти се одликују следећим особинама (*Wooldridge and Jennings, 1995; Simović, 2014*):

- Учење – способност агената да током свог рада прикупљају нова знања.
- Аутономност – способност агената да у свом раду самостално обављају постављене задатке.
- Интеракција – способност агената да комуницирају са окружењем.
- Персонализација – способност агената да се прилагођавају корисничким потребама у току рада.
- Поверљивост и мобилност – способност агената да се крећу кроз различите платформе обезбеђујући висок ниво поверљивости података.

У циљу остваривања њихове основне функције, интелигентни агенти у системима препоруке треба да процене и да предвиде који је производ за корисника вредан препоручивања. Да би то могао да уради, агент треба да буде способан да упореди карактеристике одређених производа и тада да донесе одлуку који производ ће бити препоручен.

Предикција једног неперсонализованог алгоритма за препоручивање који кориснику даје препоруку само најпопуларнијих и најбоље ранжираних производа није прецизан јер су одсутне многе потребне информације о кориснику, оценама других корисника, њихове препоруке, и информације о укупном броју корисника система који су ранжирани посматрани производ (*Simović, 2014*). У неперсонализованом систему препоруке (приказано на слици 19) нису креирани ни атрибути ни модели корисника. Систем приказује просечну оцену производа из матрице укупног броја датих оцена корисника система.



Слика 19. Неперсонализовани систем препоруке

Овако сугестивна препорука може бити ефикасна јер модератор распоређује садржај ради повећања продаје постављајући комплементарне производе непосредно близу. Такве препоруке захтевају мало усмереног напора јер купци раде евалуацију и обезбеђују ниво кредибилитета.

Аутоматски системи за персонализацију се разликују у примени метода које се користе за креирање модела корисника док се у домену алгоритамског приступа користи предвиђање. Могу да постоје различити видови помагања кориснику: (а) помоћ приликом дефинисања упита за претрагу; (б) редефинисање корисничког упита; (в) персонализација претраге; (г) поређење различитих производа или (д) других сервиса на бази одређених критеријума (Ricci et al., 2011).

Прикупљени подаци потенцијално крију важне информације које могу помоћи у разумевању ранијих активности или у оптимизацији корисничке активности у будућности. Главни проблем у великим базама података је проналажење одговарајућих узорака, праваца кретања и аномалија података, као и израда квантитативних модела из њих.

3.2.6. РЕАЛИЗАЦИЈА СИСТЕМА ПРЕПОРУКЕ

Реализован систем генерише препоруку применом колаборативног филтрирања, који се може формално записати на следећи начин (Sarwar et al., 2001; Adomavicius and Tuzhilin, 2005; Cantador et al., 2008; Simović, 2014):

ако је $U=(u_1, u_2, \dots, u_m)$ скуп корисника система, нека је $I=(i_1, i_2, \dots, i_n)$ скуп свих производа (енгл. *Items*) који се могу препоручити, а R потпуно уређени скуп где је $g:U \times I \rightarrow R$ функција корисности (енгл. *Utility Function*). Вредност функције $g(u_j, i_k)$ представља корисност производа i_k за корисника u_j . Циљ система је да за сваког корисника $u_j \in U$ одреди производ $i^{\max, u_j} \in I$ за који је вредност функције g максимална:

$$\forall u_j \in U, i^{\max, u_j} = \arg \max_{i_k \in I} g(u_j, i_k) \quad (6)$$

Корисност производа се често вреднује бројчаном оценом коју ће одредити сам корисник или ће се она израчунати. Оцена притом не мора бити вредност коју је одређени корисник доделио производу, већ се може изразити и бројем прегледа производа, бројем купљених примерака и сл. Оцена представља меру којом се исказује и предвиђа колико је за корисника одређени садржај релевантан. Да би то било могуће, сваки се корисник $u_j \in U$ описује моделом који може укључивати основне податке о кориснику, његовим преференцијама и сл. Такође, сваки производ $i_k \in I$ се описује скупом карактеристика.

Међутим, у већини случајева функција g није дефинисана на целом скупу $U \times I$, већ само на његовом подскупу. Иницијално, оцена је позната само за оне производе које су корисници претходно оценили, док се за непознате вредности она заснива на предвиђању. Према томе, један од циљева система препоруке је што тачније предвидети вредности за неоцењене производе. Након предвиђања вредности за све уређене парове скупа $U \times I$, препорука за одређеног корисника се одређује према мат. формули 6. У зависности од начина презентације препоручених садржаја, кориснику се може понудити једна или више препорука (на пример, листа од n препорука поређаних према висини оцене).

Начин на који систем за препоручивање производа методом колаборативног филтрирања настоји предвидети корисност појединог производа за активног корисника се базира на основу вредновања производа од стране осталих корисника система. Вредност функције $g(u_j, i_k)$ којом се представља корисност производа $i_k \in I$ за корисника $u_j \in U$ се у том случају процењује и заснива на познатим вредностима $g(u_s, i_k)$ за производ i_k при чему је корисник $u_s \in U$ сличан кориснику $u_k \in U$.

Дефинише се модел корисника, у ознаци $CollaborativeUserProfile(u_m)$ као вектор чије су компоненте оцене које је корисник доделио доступним производима у систему:

$$CollaborativeUserProfile(u_m) = r_m = (r_{m,1}, r_{m,2}, \dots, r_{m,n}) \in R^N$$

Оцене које су одређеном производу доделили остали корисници представљају се вектором $Ratings(i_n)$:

$$Ratings(i_n) = r_n = (r_{1,n}, r_{2,n}, \dots, r_{M,n}) \in R^M$$

У случају да корисник u_i није оценио садржај i_j за компоненту $r_{i,j}$ у оба горе наведена вектора вреди вредност $r_{i,j} = \emptyset$. Вредност корисности производа $i_n \in I$ за корисника $u_m \in U$ одређује се функцијом $score$ која при израчунавању комбинује компоненте модела корисника u_m и оцене за i_n :

$$g(u_m, i_n) = score(CollaborativeUserProfile(u_m), Ratings(i_n)) \in R$$

Колаборативно филтрирање упоређује извршену сваку куповину оствареног купца, рангира производе по сличности, и комбинацијом сличних производа генерише препоруку. Алгоритам генерише табелу сличности како би утврдио производ који је

најприкладнији кориснику на основу његовог задатог упита. Начин на који алгоритам израчунава сличности између посматраног производа са свим другим производима електронског каталога је исказан у следећем псеудо коду (*Linden et al., 2003; Simović, 2014*):

```

For each item in product catalog,  $I_1$ 
  For each customer  $C$  who purchased  $I_1$ 
    For each item  $I_2$  purchased by
      customer  $C$ 
        Record that a customer purchased  $I_1$ 
          and  $I_2$ 
  For each item  $I_2$ 
    Compute the similarity between  $I_1$  and  $I_2$ 

```

Могуће је израчунавање сличности између два производа на различите начине. Уобичајен метод је косинус (енгл. *Cosine Similarity*) израчунавања где сваки вектор кореспондира са производом а не са купцем, док вектор M -димензије кореспондира са купцем који је обавио куповину посматраног производа.

Када је реч о електронским продавницама великих складишта података, овакав начин израчунавања за генерисање табеле сличних производа захтева време са случајем $O(N^2M)$. Међутим, у пракси је ближе случају $O(NM)$ јер многи купци имају мањи број остварених куповина.

Алгоритам за колаборативно филтрирање приказује купца са N -димензионалним вектором производа, где је N број производа електронског каталога. Компоненте вектора су позитивне за обављене куповине или позитивно оцењене, док су негативне компоненте вектора негативно оцењене. Компензацијом, ради најбоље продаје производа електронског каталога, алгоритам умножава компоненте вектора инвертујући фреквенцију (инвертује број купаца који су купили или оценили производ) стварајући мање познате производе више релевантним. Скоро за све купце, приликом одабира производа за куповину, приказ препорука коришћењем овог вектора је редак. Алгоритам генерише препоруку базирану на само неколико купаца сличних преференција у односу на посматраног корисника.

Преференције и сличности два купца, A и B , систем може мерити на различите начине. Примењен метод је израчунавање косинуса угла између два вектора (мат. формула 7) (*Linden et al., 2003; Simović, 2014*):

$$similarity(\vec{A}, \vec{B}) = \cos(\vec{A}, \vec{B}) = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\| * \|\vec{B}\|} \quad (7)$$

Систем препоручује производе на основу сличности коришћењем различитих метода. Уобичајена техника генерисања препоруке је рангирање сваког производа електронског каталога на основу тога колико је купаца купило исти производ.

Ако је $O(MN)$, где је M број остварених купаца и N број производа у електронском каталогу, систем истражује M купаца до N производа за сваког купца. Међутим, из разлога што је просечан вектор купца изузетно редак, перформанса алгоритма система тежи близини $O(M+N)$.

Анализа сваког оствареног купца је приближна $O(M)$, а не $O(MN)$. Разлог томе је што вектор скоро сваког купца садржи мали број производа без обзира на величину података у е-каталогу. Међутим, постоји мали број купаца који су обавили куповину или оценили мали проценат производа из каталога захтевајући $O(N)$ време извршавања, тако да је финална перформанса алгоритма приближна $O(M+N)$.

Код електронских продавница које складиште велике количине података, нпр. милион и више купаца и исто толико или више производа у каталогу, приказани алгоритам система би се сусрео са потешкоћама као што је брзина очекиваних калкулација и израчунавања, скалирања и укупних перформанси система.

Могуће је делимично решење проблема скалирања смањивањем количине ускладиштених података уколико је база преоптерећена, што важи за велике електронске продавнице са милионским записима. M се може смањити случајним узорком купаца из базе, одбацавањем купаца који су извршили мањи број куповина, док се N може смањити одбацавањем или веома популарних или непопуларних производа. Коришћење техника за смањивање, као што је *clustering*, резултује умањивањем елемената M и N великим фактором.

Међутим, понуђено решење проблема и поменуте технике за смањивање базе утичу на квалитет система препоруке на много начина. Алгоритам истражује само мали узорак купаца тако да посматраном кориснику систем неће дати препоруке садржаја од интереса. Ако алгоритам одбаци или најпопуларније или непопуларне производе, они се више неће појављивати као препоруке, а купци који су обавили куповине тих производа неће добити адекватне препоруке јер је генерисана листа умањена.

Реализован систем препоруке са приказом генерисане листе приликом избора жељеног производа за куповину је приказан на слици 20.

Elektronska trgovina – Priručnik
[350] 700RSD

Autori: prof. dr Miroslav D. Lutovac i mast. inž. Aleksandar Simović

Sadržaj:

- Imperativ elektronske trgovine
- Instalacija Apache Web servera
- Instalacija CMS OpenSource softvera osCommerce
- Izlog elektronske prodavnice
- Pasusi na početnoj strani
- Konfiguracija elektronske prodavnice
- Online catalog
- Registracija kupaca, kupovina & izveštavanje
- osCommerce Simple Template System
- Info boksovi
- Bezbednost

Изабран производ за куповину

Препоручена литература на основу изабраног производа

ISBN: 978-86-7982-116-4

Komentari Dodaj u korpu

Korisnici koji su kupili ovaj udžbenik, kupili su i:

Osnovi računarske tehnike – Zbirka zadataka	Arhitektura i organizacija računara 1 – Priručnik	Zbirka rešenih zadataka iz telekomunikacija

Слика 20. Реализован систем препоруке са приказом генерисане листе

3.2.7. КОМПАРАТИВНА АНАЛИЗА СИСТЕМА ПРЕПОРУКА

У компаративној анализи система препоруке колаборативног филтрирања за системе филтрације базиране на корисницима и системе филтрације базиране на производима коришћен је софтвер отвореног кода *Apache Mahout*, проистекао као сервис пројекта *Hadoop*.

Велике библиотеке *Hadoop* екосистема омогућавају дистрибуирану обраду и процесирање велике запремине података обрађујући их преко кластера рачунара користећи програмерске моделе. Библиотеке су пројектоване тако да детектују и решавају могуће грешке на апликационом слоју ради доступности сервиса.

Apache Mahout омогућава својим компонентама изградњу и конструисање система препоруке избором одговарајућих алгоритама. Софтвер садржи модуле са следећим кључним апстракцијама (*Owen and Owen, 2012; Bagchi, 2015*):

- Модел података (енгл. *Data Model*);
- Сличности корисника у систему (енгл. *User Similarity*);
- Сличности производа у систему (енгл. *Item Similarity*);
- Систем препоруке (енгл. *Recommender System*).

Циљ пројекта *Apache Mahout* је изградња окружења за креирање скалабилне, прилагодљиве апликације машинског учења. *Mahout* библиотеке омогућавају опсежне анализе и евалуације прилагодљивих алгоритама машинског учења (*Holmes, 2012*). Модел машинског учења је примењен у анализи и евалуацији колаборативног филтрирања. За развој и имплементацију модела за машинско учење неопходно је утврдити критеријуме прикупљања података. Скуп података треба да садржи потребне информације које могу допринети већој тачности израчунавања. Архитектура система машинског учења треба да обухвати изворе података, апликације и алате за машинско учење као и механизам за интеграцију са постојећим системом (*Simović, 2017*).

За анализу колаборативног филтрирања коришћен је метод израчунавања коефицијента Пирсонове корелације – мера степена линеарне релације која постоји између две варијабле (приказано на мат. формули 5). Пирсонов коефицијент корелације је изведен из модела линеарне регресије која почива на претпоставци да, независно од података релација мора бити линеарна; све грешке морају бити независне; и константну варијацију за свако подешавање независних варијабли. Када те претпоставке нису задовољене, Пирсонов коефицијент корелације не показује индикације тачности у израчунавању (*Herlocker et al., 1999*).

У анализи колаборативног филтрирања по корисницима (део референтног кода је приказан на слици 21) коришћена је евиденција сервера електронске продавнице књига која садржи ~3900 вредновања; ~450 корисника и ~1495 производа који су узети за израчунавање.

У анализи система важан чинилац је покривеност. Покривеност је проценат од укупног броја производа за који ће систем препоруке обезбедити предвиђање. Основно мерење покривености је проценат доступних производа за предвиђање. Заједничке карактеристике система које могу да смање покривеност су мале димензије сличности корисника и њихово узорковање ради проналажења блискости. Покривеност подразумева проценат свих корисника система за које се предвиђање тражи (*Simović, 2018*).

Експериментални резултати анализе имају високу покривеност која износи око 95%. Максимална покривеност није обезбеђена из следећих разлога: (а) постоје одређени производи без вредновања (купљени/оцењени); (б) мали број корисника који је вредновао одређени производ; (ц) мала сличност појединих корисника система (*Simović, 2018*).

```

package rs.prodavnicaknjiga.FiltriranjePoKorisnicima;

import java.io.File;
import java.util.List;

import org.apache.mahout.cf.taste.impl.model.file.FileDataModel;
import org.apache.mahout.cf.taste.impl.neighborhood.ThresholdUserNeighborhood;
import org.apache.mahout.cf.taste.impl.recommender.GenericUserBasedRecommender;
import org.apache.mahout.cf.taste.impl.similarity.PearsonCorrelationSimilarity;
import org.apache.mahout.cf.taste.model.DataModel;
import org.apache.mahout.cf.taste.neighborhood.UserNeighborhood;
import org.apache.mahout.cf.taste.recommender.RecommendedItem;
import org.apache.mahout.cf.taste.recommender.UserBasedRecommender;
import org.apache.mahout.cf.taste.similarity.UserSimilarity;

/** www.prodavnicaknjiga.rs */

public class App
{
    public static void main( String[] args ) throws Exception
    {
        DataModel model = new FileDataModel(new File("Podaci/logfile.csv"));
        UserSimilarity similarity = new PearsonCorrelationSimilarity(model);
        UserNeighborhood neighborhood = new ThresholdUserNeighborhood(0.1, similarity, model);
        UserBasedRecommender recommender =
            new GenericUserBasedRecommender(model, neighborhood, similarity);
        List<RecommendedItem> recommendations = recommender.recommend(350, 3);
        for (RecommendedItem recommendation : recommendations) {
            System.out.println(recommendation);
        }
    }
}

```

Слика 21. Колаборативно филтрирање по корисницима – део референтног кода

Линија кода (црвени оквир на слици 21):

```
List<RecommendedItem> recommendations = recommender.recommend(350, 3);
```

генерише препоруке за активног корисника система за посматрани производ са шифром 350. Систем приликом генерисања листе узима у разматрање све кориснике и сва вредновања посматраног производа; креира групу корисника сличних интересовања а излазни резултат одговара преференцијама посматраног корисника са групом најближих у систему.

Израчунавање се базира на претпоставци да ће кориснику који је изабрао производ са шифром 350 одговорати производи који су остали корисници система вредновали највишим оценама (купили/оценили).

Резултат израчунавања је приказан на слици 22.

```

RecommendedItem[item:1302, value:5.0]
RecommendedItem[item:1367, value:5.0]
RecommendedItem[item:1237, value:4.7962523]

```

Слика 22. Генерисана препорука колаборативног филтрирања по корисницима (прва колона: производ 350, друга колона: препоручен производ, трећа колона: рејтинг производа од стране осталих корисника у систему при чему је 5.0 највиша оцена)

За следеће израчунавање колаборативног филтрирања базираног на производима коришћена је иста евиденција сервера електронске продавнице књига.

Део референтног кода колаборативног филтрирања по производима у систему препоруке је приказан на слици 23.

Резултат израчунавања је приказан на слици 24.

У израчунавању, линија кода (црвени оквир на слици 23):

```
List<RecommendedItem>recommendations = recommender.mostSimilarItems(itemId, 3);
```

прави петљу кроз електронски каталог, пореди карактеристике свих производа у систему и генерише препоруку за три најближа производа према садржају.

```
package rs.prodavnicaknjiga.FiltriranjePoProizvodima;

import java.io.File;
import java.io.IOException;
import java.util.List;

import org.apache.mahout.cf.taste.common.TasteException;
import org.apache.mahout.cf.taste.impl.common.LongPrimitiveIterator;
import org.apache.mahout.cf.taste.impl.model.file.FileDataModel;
import org.apache.mahout.cf.taste.impl.recommender.GenericItemBasedRecommender;
import org.apache.mahout.cf.taste.model.DataModel;
import org.apache.mahout.cf.taste.recommender.RecommendedItem;
import org.apache.mahout.cf.taste.similarity.ItemSimilarity;

public class FiltriranjePoProizvodima {

    public static void main(String[] args) {
        try {

            DataModel dm = new FileDataModel(new File("Podaci/logfile.csv"));
            GenericItemBasedRecommender recommender = new GenericItemBasedRecommender(dm, sim);
            int x=1;
            for(LongPrimitiveIterator items = dm.getItemIDs(); items.hasNext();) {
                long itemId = items.nextLong();
                List<RecommendedItem>recommendations = recommender.mostSimilarItems(itemId, 3);
                for(RecommendedItem recommendation : recommendations) {
                    System.out.println(itemId + "," + recommendation.getItemID() + "," + recommendation.getValue());
                }
            }
        }
    }
}
```

Слика 23. Колаборативно филтрирање по производима – део референтног кода

349,908,0.21428572
349,898,0.1891892
350,752,0.21212122
350,351,0.19607843
350,355,0.1884058
351,350,0.19607843
351,349,0.18604651
351,908,0.1764706

Слика 24. Генерисана препорука колаборативног филтрирања по производима (вредности поклапања и сличност производа у систему – прва колона: производ, друга колона: сличан производ, трећа колона: вредност поклапања при чему је 1 највеће поклапање)

У колаборативном филтрирању постоје многе анализе тачности мерења и евалуације система. Суштински се деле у две главне категорије: 1) статистичка мерења тачности (енгл. *Statistical Accuracy Metrics*) и 2) тачност система у доношењу одлука (енгл. *Decision Support Accuracy Metrics*) (Melville et al., 2002; Herlocker et al., 1999).

Прва категорија мерења процењује и израчунава тачност колаборативног филтрирања упоређујући нумеричке вредности предикције у односу на вредновање производа од стране осталих корисника система. Друга категорија израчунава ефикасност предвиђања, односно колико предвиђање опредељује корисника да бира високо рангиране производе из препоруке.

Статистичка процена тачности система на примењеном скупу података коришћењем *Mahout* оквира, извршена је следећом линијом кода (зелени оквир на слици 25):

```
org.apache.mahout.cf.taste.eval.RecommenderEvaluator
```

За евалуацију система препоруке, сличности и процене статистичке тачности је коришћен метод израчунавања коефицијента Пирсонове корелације колаборативног филтрирања по корисницима (плави оквир на слици 25):

```
org.apache.mahout.cf.taste.impl.similarity.PearsonCorrelationSimilarity;
```

са моделом израчунавања апсолутне просечне различитости (црвени оквир на слици 25):

```
RecommenderEvaluator evaluator = new AverageAbsoluteDifferenceRecommender  
Evaluator();
```

од изабраног броја корисника од $n/10$ (жути оквир на слици 25) до $n/50$ (зелени оквир на слици 26) од укупног броја корисника у систему узетих за процену.

```

aluacijaSistemaFiltriranjaPoKorisnicima.java ☒
package rs.prodavnicaknjiga.FiltriranjePoKorisnicima;

import java.io.File;
import java.io.IOException;

import org.apache.mahout.cf.taste.common.TasteException;
import org.apache.mahout.cf.taste.eval.RecommenderBuilder;
import org.apache.mahout.cf.taste.eval.RecommenderEvaluator;
import org.apache.mahout.cf.taste.impl.eval.AverageAbsoluteDifferenceRecommenderEvaluator;
import org.apache.mahout.cf.taste.impl.model.file.FileDataModel;
import org.apache.mahout.cf.taste.impl.neighborhood.ThresholdUserNeighborhood;
import org.apache.mahout.cf.taste.impl.recommender.GenericUserBasedRecommender;
import org.apache.mahout.cf.taste.impl.similarity.PearsonCorrelationSimilarity;
import org.apache.mahout.cf.taste.model.DataModel;
import org.apache.mahout.cf.taste.neighborhood.UserNeighborhood;
import org.apache.mahout.cf.taste.recommender.Recommender;
import org.apache.mahout.cf.taste.similarity.UserSimilarity;

public class EvaluacijaSistemaFiltriranjaPoKorisnicima {

    public static void main(String[] args) throws Exception {
        DataModel model = new FileDataModel(new File("Podaci/logfile.csv"));
        RecommenderEvaluator evaluator = new AverageAbsoluteDifferenceRecommenderEvaluator();
        RecommenderBuilder builder = new MyRecommenderBuilder();
        double result = evaluator.evaluate(builder, null, model, 0.9, 1.0);
        System.out.println(result);
    }
}

```

Слика 25. Евалуација система базираног на филтрирању по корисницима – део референтног кода за $n/10$ од укупног броја корисника система

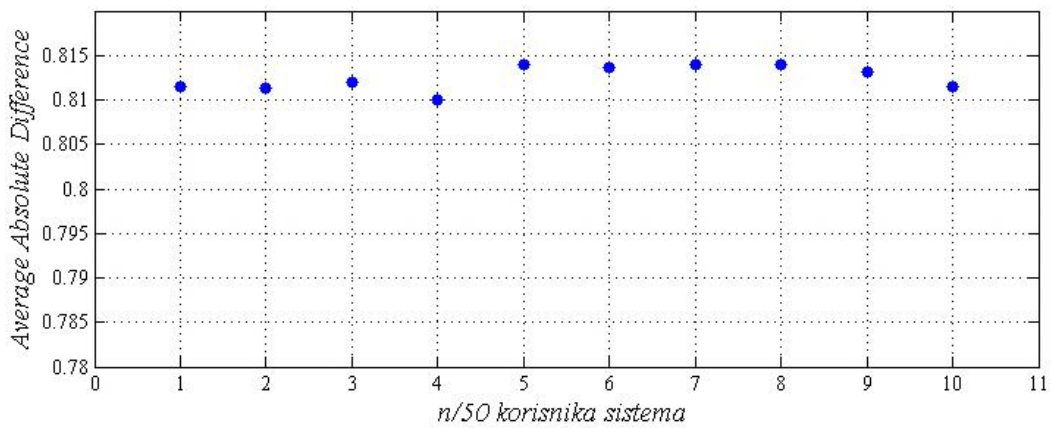
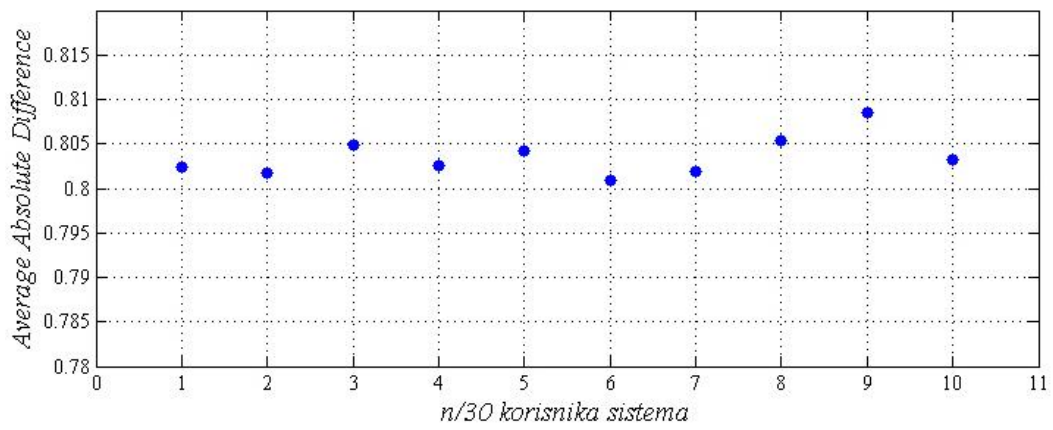
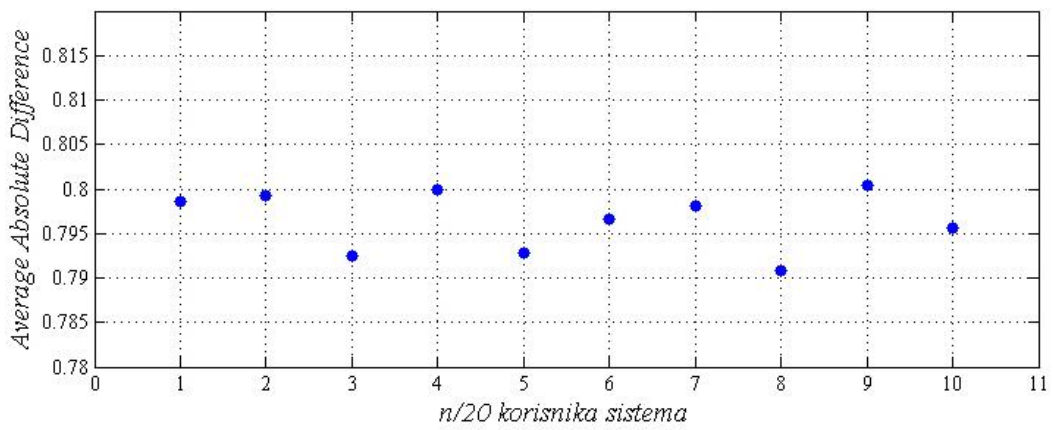
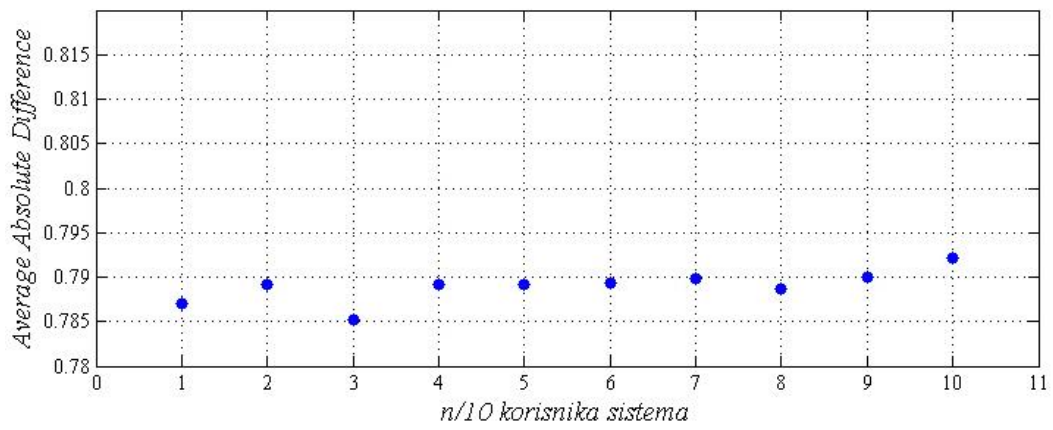
```

public class EvaluacijaSistemaFiltriranjaPoKorisnicima {

    public static void main(String[] args) throws Exception {
        DataModel model = new FileDataModel(new File("Podaci/logfile.csv"));
        RecommenderEvaluator evaluator = new AverageAbsoluteDifferenceRecommenderEvaluator();
        RecommenderBuilder builder = new MyRecommenderBuilder();
        double result = evaluator.evaluate(builder, null, model, 0.5, 1.0);
        System.out.println(result);
    }
}

```

Слика 26. Евалуација система базираног на филтрирању по корисницима – део референтног кода за $n/50$ од укупног броја корисника система



Слика 27. Визуализација добијених резултата израчунавања

Експериментални резултат евалуације система колаборативног филтрирања по корисницима је показао да се са повећањем броја корисника у систему тачност препоруке смањује и уколико је *AbsoluteDifference* приказ мањи, утолико је тачност система већа (приказано на слици 27).

Постоје могућа ограничења која се односе на ову компаративну анализу. Прво, скуп података који је коришћен за примену решења за анализу предвиђања је један скуп који је интегрисан у представљени оквир. Друго, у циљу успешне анализе и увида у знање како би се издвојила вредност, организације треба да имплементирају одговарајући *big data* оквир који може да обради више независних извора различитих података. У том редоследу, ако је већа количина и разноврсност скупа података који је припремљен за обраду, постоје веће могућности за проналазак скривених увида у обрасце. Међутим, лоше идентификовани улазни подаци резултоваће лошим квалитетом без обзира колико је исправно *big data* решење (*Simović, 2018*).

Да би постигли конкурентску предност у *big data* окружењу, организације треба да пројектују *big data* оквир са сервисима и модулима од значаја који ће допринети задовољству корисника и дугорочном поверењу. Инфраструктура таквог система има способност да управља и релационим и нерелационим структурама података и да извршава обраду у реалном времену коришћењем стриминг алата и модула за интерну интероперабилност са спољним изворима. Исто тако, има способност за анализу предвиђања и интерактивно истраживање обједињених података из различитих извора (*Simović, 2017*).

4. МОДЕЛ ПАМЕТНЕ БИБЛИОТЕКЕ ЗАСНОВАН НА *BIG DATA* ТЕХНОЛОГИЈАМА

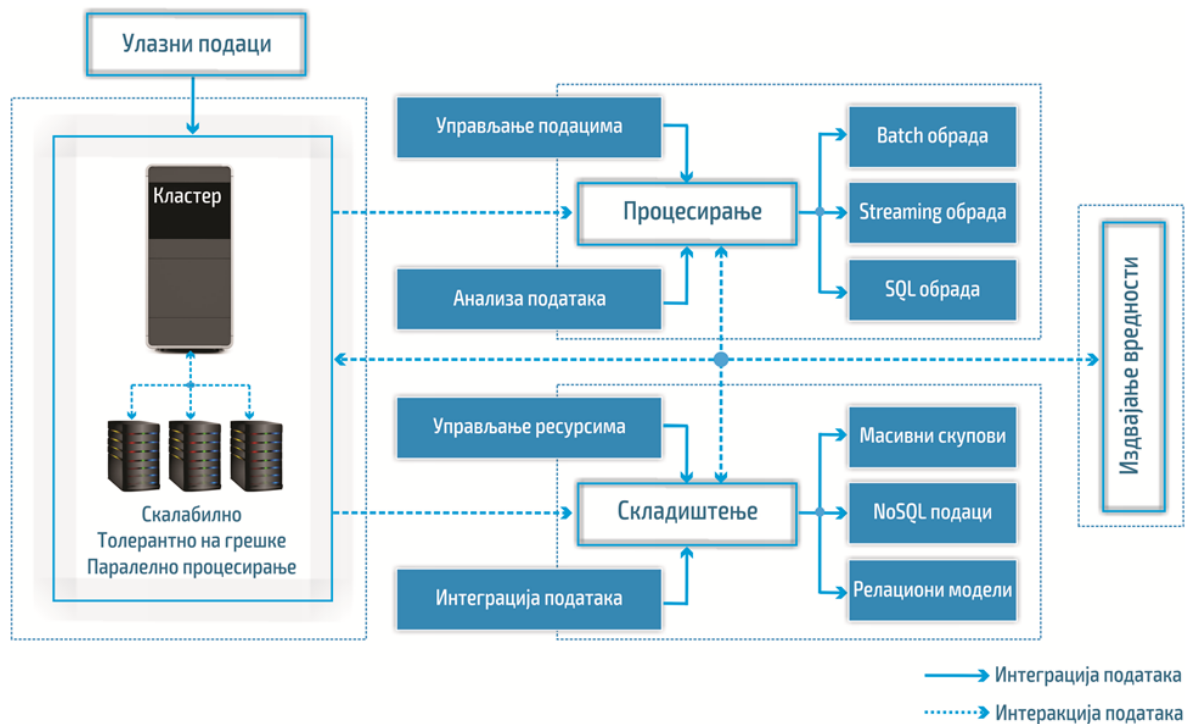
Модел паметне библиотеке треба да обухвати интеграцију *big data* технологија и система препоруке у циљу изградње система који може препоручити персонализовани садржај са повећаном прецизношћу анализом корисничких интереса прикупљених из више извора, као и карактеристике садржаја из различитих врста података.

Овај процес треба да повећа квалитет сценарија препоруке јер посматра однос између библиотеке и образовања као нераздвојне једне од друге (Simović, 2018).

Модел паметне библиотеке заснован на *big data* технологијама треба да обезбеди четири кључне карактеристике (приказано на слици 28):

- **скалабилну и еластичну инфраструктуру.** Традиционално конфигурисани системи и релационе базе података налазе се на једном серверу и могу да скалирају додавањем веће процесорске снаге, више меморије и више екстерних складишта. Као један од основних принципа анализе и складиштења података у *big data* окружењу, модел паметне библиотеке који је заснован на *big data* технологијама треба да обезбеди:
 - a) скалабилност високог нивоа са одликама паралелног процесирања, малог кашњења и снагом еластичног повећања базе; и
 - b) изградњу архитектуре кластера са паралелним опоравком и репликацијом информација, минимизирајући вероватноћу квара нодова рачунара који садрже дистрибуиране податке;
- **интеграцију сложених и разноврсних података.** Архитектура модела паметне библиотеке треба да омогући:
 - a) проналажење, преузимање, процесирање и складиштење структурираних, неструктурираних и масовних података обезбеђујући њихову кохерентност, тачност и поузданост; и
 - b) покретање послова који се могу изводити без интеракције крајњег корисника;
- **интероперабилност независних извора.** Екосистем модела треба да буде свеprisутан (енгл. *Ubiquitous*) и да укључи интеграцију и повезивање више различитих извора независних података, апликативних процеса и *big data* технолошких система, притом обезбеђујући:
 - a) паметно окружење за размену информација и колаборацију; и

- b) сложену анализу међусобно повезаних послова како би пружили нове увиде за одлуке и акције које побољшавају пословне исходе и искуство крајњег корисника;
- **издвајање вредности комбиновањем различитих типова података.** *Big data* технологије треба да издвоје и створе нова знања јачајући доступност информација анализом обједињених података. Библиотечки информациони систем, као саставни део информационог система образовне институције, и као технолошка и интелектуална инфраструктура, треба да омогући прикупљање података, приступ информацијама и производњу знања са циљем стварања културне и научне вредности за заједницу као основне премисе свог постојања.



Слика 28. Кључне карактеристике модела паметне библиотеке заснованог на *big data* технологијама

4.1. ПРОЈЕКТОВАЊЕ АРХИТЕКТУРЕ МОДЕЛА ПАМЕТНЕ БИБЛИОТЕКЕ ЗАСНОВАНОГ НА *BIG DATA* ТЕХНОЛОГИЈАМА

Као један од основних фокуса паметне библиотеке је интеграција различитих података из више различитих извора и пружање прилагођених препорука кориснику са циљем реализације прецизнијег и ефикаснијег система препоруке (приказано на слици 29).

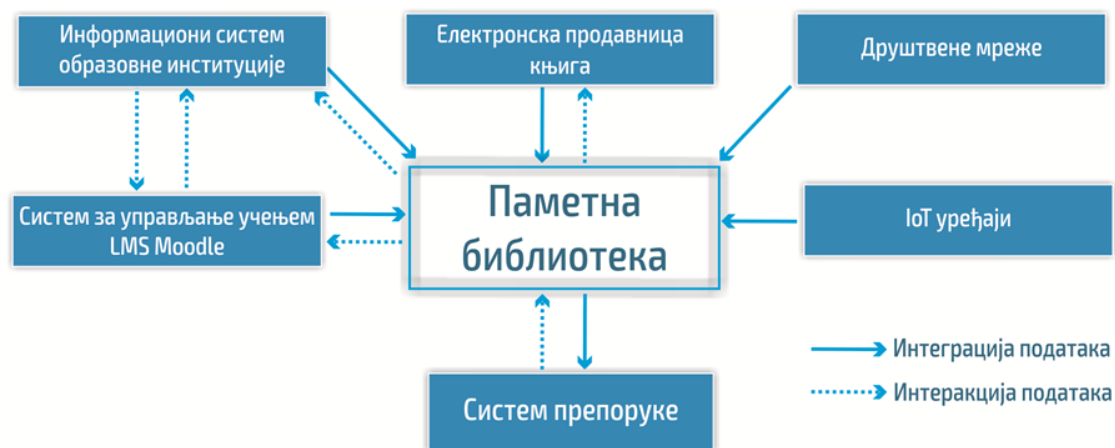
Интеракција између разноврсних података интегрисаних из више извора приликом генерисања персонализованих препорука је представљена на иновативан начин уз

приступачан систем који постиже највећу интероперабилност.

У предложеном моделу подаци се сакупљају и интегришу у паметну библиотеку из више извора. Архитектура система паметне библиотеке обезбеђује интеграцију и интеракцију података прикупљених:

- из информационог система образовне институције – евиденције о генерисаним моделима студената, као корисника који укључују биране предмете у текућој школској години и положене предмете са просечном оценом из претходних година студија;
- са сервера електронске продавнице књига образовне институције – евиденције о преференцијама студената и других корисника, које укључују купљену и оцењену стручну литературу;
- са *LMS Moodle* платформе – информације које студенти размењују између себе и наставника путем форума за вести, укључујући одабране курсеве студената током циклуса учења;
- са друштвених медија – архивирани поруке са друштвених медија које изражавају мишљење студената у виду оцене квалитета књиге или аутора према субјективном мишљењу и задовољству студента као корисника;
- са *IoT* уређаја – који обезбеђују управљање процесом задуживања и аутоматском контролом доступности књига у виду информација о локацијама штампаних издања књига у библиотеци.

Ови подаци се прикупљају, обрађују и анализирају у *big data* екосистему ради реализације и визуализације коначног садржаја – персонализоване препоруке садржаја од већег интереса на основу корисничких потреба.



Слика 29. Блок дијаграм архитектуре система паметне библиотеке

4.1.1. СТРУКТУРА МОДЕЛА ПАМЕТНЕ БИБЛИОТЕКЕ

Структура предложеног модела паметне библиотеке заснованог на *big data* технологијама, са међусобним везама између компонената и поткомпонената је приказана на слици 30.

Прва компонента је систем паметне библиотеке у оквиру које су предвиђене поткомпоненте; док друга компонента обухвата интеграцију података из више различитих извора који се интегришу у паметни екосистем, и над којима се извршава обрада, процесирање, анализа, визуализација, складиштење и интеракција, са коначним циљем издвајања вредности у виду система препоруке од већег интереса за корисника из прикупљених и ускладиштених података.

Модел паметне библиотеке чине следеће компоненте:

Компонента 1 – К1: Систем паметне библиотеке. Ова компонента обухвата следеће поткомпоненте:

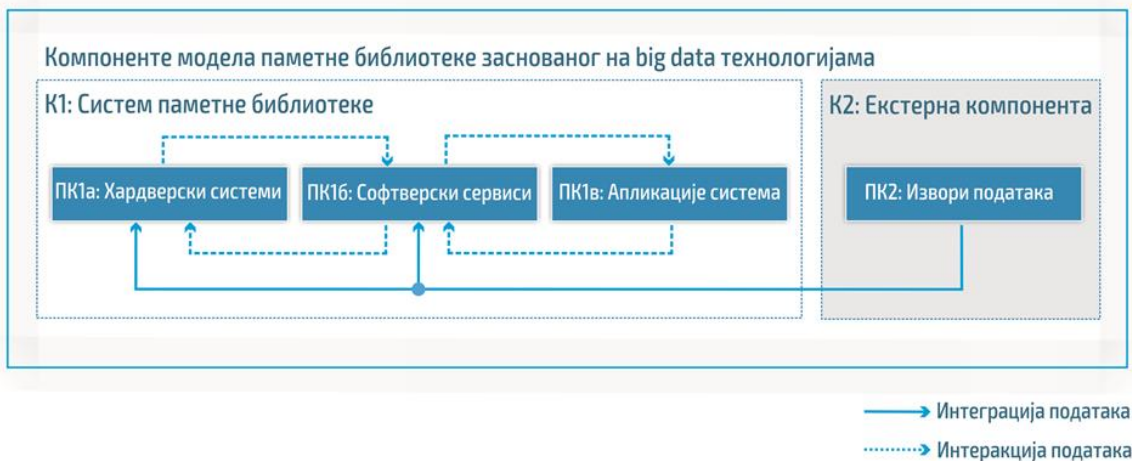
ПК1а: Хардверски системи – треба да омогуће изградњу инфраструктуре која је способна да лоцира, прикупи и ускладишти потребне податке обезбеђујући њихову кохерентност.

ПК1б: Софтверски сервиси – треба да омогуће обраду, анализу, процесирање, визуализацију, интеграцију и интеракцију прикупљених и ускладиштених података.

ПК1в: Апликације система – треба да омогуће издвајање вредности и информације од значаја из прикупљених података ради генерисања листе препорука од већег интереса за посматраног корисника паметног библиотечког информационог система.

Компонента 2 – К2: Екстерна компонента у оквиру које је предвиђена поткомпонента:

ПК2: Извори података. Ова компонента обухвата интеграцију различитих података из више различитих извора који садрже информације које носе карактеристике од значаја ради генерисања препоруке садржаја.



Слика 30. Компоненте модела паметне библиотеке заснованог на big data технологијама – приказ структуре модела са међусобним релацијама

4.1.2. ПРОЈЕКТОВАЊЕ ХАРДВЕРСКОГ СИСТЕМА ПАМЕТНЕ БИБЛИОТЕКЕ

Пројектовање предложеног хардверског система је подстакнуто запажањима у оптерећењу постојећих традиционалних информационих система образовних институција и технолошким окружењем, тренутним и очекиваним, која треба да одражавају изразит одмак од ранијих традиционалних избора у пројектовању информационих инфраструктура образовних институција.

Прво, кварови компонената су норма, а не изузетак. Системи датотека могу да се састоје од неколико стотина рачунара предвиђених за управљање складиштем података и који су направљени од повољних / јефтних хардверских компоненти чији квалитет практично не гарантује њихову поузданост, нити опоравка од кварова ни у једном тренутку. Проблеми се такође идентификују због грешака у оперативном систему, апликацијама, људских грешака, као и кварова дискова, меморије и других хардверских делова. Из тих разлога, константно и континуирано надгледање система, благовремено откривање грешака, толеранција на грешке (енгл. *Fault tolerance*) и аутоматски опоравак морају бити неизбежан саставни део система.

Друго, скупови података у *big data* окружењу су према традиционалним стандардима, огромни. Сваки од скупова великих података сачињава мноштво датотека од којих свака датотека садржи много објеката апликација, нпр. као што су веб документи. Када се за циљ задаје успешно управљање брзорастућим набором информација изражених у терабајтима који садрже милијарде објеката, претпоставке традиционалних приступа у пројектовању информационих структура, укупна величина блокова података и улазно / излазни параметри дизајна, морају бити ревидирани.

У складу са наведеним, пројектовање хардверске инфраструктуре система паметне библиотеке треба да буде вођен, како изазовима, тако и могућностима:

- континуираног надгледања са способношћу детектовања грешака, толеранције и опоравка на грешке на рутинској основи;
- складиштења велике количине различитих података;
- ефикасног управљања ускладиштеним подацима;
- високих перформанси клијентске апликације, пројектоване преко улазно / излазних, до издвојених и лоцираних вредности;
- имплементације и конфигурације добро дефинисаних нодова рачунара и сервера кластера који истовремено придружују датотеке, редувантно чувајући њихове реплике;
- децентрализације сервера делегирањем надлежности нодова кластера, анулирајући уско грло примарног *Master Node*-а.

Хардверски систем предложеног модела паметне библиотеке се састоји из три главне категорије хардверских улога у имплементацији и конфигурацији кластера: (1) клијент, (2) *Master Node* и (3) *Slave Node*-ови.

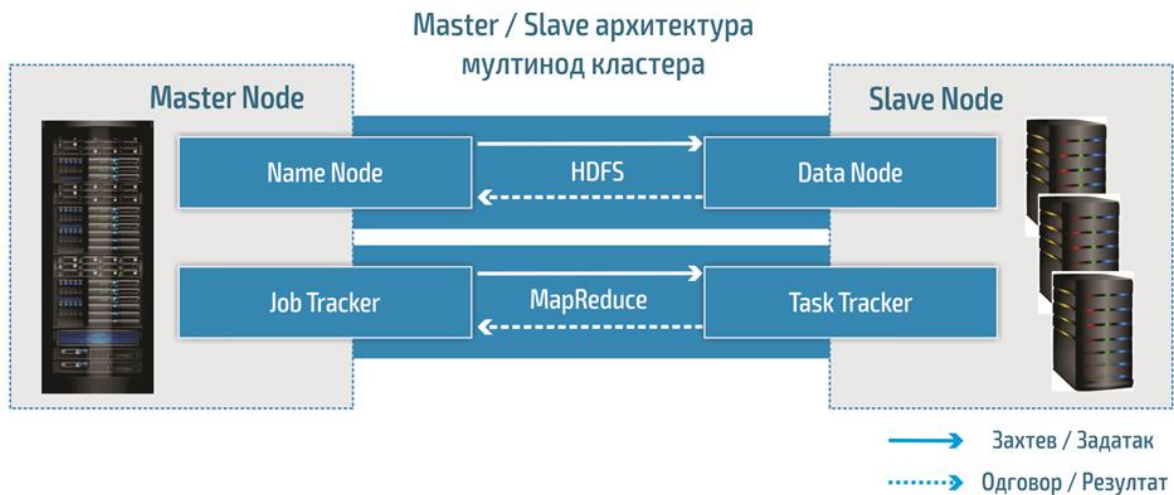
Кластер чини један примарни *Master Node* и вишеструки нод сервери за обављање задатих послова којима може приступити више клијената. Сервери у кластеру раде заједно како би обезбедили високу доступност, поузданост и скалабилност у погледу пружања услуга клијентима са континуираним надгледањем дистрибуираних реплика и метаподацима о сачуваним датотекама. Ако један рачунар / сервер не успе у извршавању задатих послова, ресурси се дистрибуирају између осталих машина у кластеру.

Задаци за извршавање могу бити покренути, како у кластеру са једним нодом (енгл. *Single Node Cluster*), тако и у мултинод кластеру (енгл. *Multinode Cluster*). Кластер са једним нодом или псеудо-дистрибуирани кластер (енгл. *Pseudo-distributed Cluster*) је онај у којем сви нодови попут *Name Node*-а, *Data Node*-а, *Job Tracker*-а, и *Task Tracker*-а раде на једној машини.

У пројектовању предложеног система паметне библиотеке примењен је мултинод кластер који је у својој основи конфигурисан на следећи начин (приказано на слици 31):

- *Master Node* и *Slave Node*-ови раде на различитим машинама;
- *Master Node* покреће *Name Node* и *Job Tracker*;
- *Slave Node* покреће *Data Node*-ове у кластеру и *Task Tracker*;

- Сваки *Slave Node* добија инструкције од *Master Node*-а;
- *Master Node* извршава две кључне функционалне карактеристике које дефинишу кластер: (1) чува податке преко *HDFS* (енгл. *Hadoop Distributed File System*) слоја софтверског оквира; и (2) извршава паралелна израчунавања из дистрибуираних података преко *MapReduce* слоја софтверског оквира (детаљан опис *HDFS* и *MapReduce* процеса у предложеном моделу паметне библиотеке је представљен у следећем поглављу);
- *Name Node* надгледа и координира функцију складиштења података на *HDFS*-у;
- *Job Tracker* надгледа и координира паралелну обраду података применом *MapReduce* процеса;
- *TaskTracker* извршава захтеве / задатке у зависности од (енгл. *Slave to*) *JobTracker*-а;
- *Data Node*-ови у кластеру извршавају захтеве / задатке у зависности од (енгл. *Slave to*) *Name Node*-а.



Слика 31. Архитектура предложеног хардверског система паметне библиотеке – *Master / Slave* мултинод кластер – Основна конфигурација система

Међутим, иако псеудо-дистрибуирани кластер, где исти нод делује као један мастер, може значајно поједноставити конфигурацију инфраструктуре, у извршавању задатака читања и писања може постати уско грло јер клијенти готово никада не читају и не уписују податке преко *Master Node*-а, већ од њега траже информацију који сервери имају реплика блокава који чине дистрибуирану датотеку. Уместо тога, хардверска архитектура треба да буде конфигурисана тако да клијент извршава упит *Master Node*-у о простору имена датотеке и индексу свих реплика дистрибуираних блокава података.

У пројектовању хардверске архитектуре система паметне библиотеке, мултинод кластер је конфигуриран тако да примарни задатак *Master Node*-а чини надгледање и управљање са три кључне врсте метаподатака:

- простором имена датотека и дистрибуираних блокова;
- мапираним блоковима података који чине дистрибуирану датотеку;
- локацијом реплика дистрибуираних блокова података.

Детаљан приказ пројектоване *Master / Slave* архитектуре мултинод кластера је приказан на слици 32 и описан је у наставку текста.

Name Node Master сервера управља и надгледа операције извршене на кластеру и регулише приступ датотекама од стране клијента. Приликом покретања система, *Master Node* добија информације од *Name Node*-а о локацији реплика дистрибуираних блокова података, укључујући информације које *Data Node*-ови похрањују о репликама датог блока, након чега ће моћи да контролише све позиције блокова података датотеке надгледајући статус *Name Node*-а. Такође, важно је напоменути, да приликом конфигурације кластера, улога *Master Node*-у о континуираном чувању података о локацији свих реплика може бити додељена, међутим у том случају појавио би се проблем синхронизације *Master Node*-а са *Name Node*-ом када се нови *Data Node*-ови придружују кластеру, постојећи напуштају, мењају име или поново покрећу. У кластеру са више стотина рачунара и сервера овакви догађаји су честа појава (*Ghemawat et al.*, 2003).

Name Node – Одржава хијерархију простора имена и метаподатке фајл система као што су локације блокова података (три реплике сваког блока у кластеру са конфигурационим `dfs.replication` параметром), где уобичајена величина једног блока износи 64 мегабајта. Као и у фајл систему за један диск, датотеке су у кластеру раздвојене на блокове који се чувају као независне јединице. Апстракције блокова података за дистрибуирани фајл систем доноси многе предности. Примера ради, ако је датотека која је спремна за обраду већа од било којег појединачног диска на серверу, а притом не постоје специфичне директиве да се блокови чувају на истом диску, један блок може бити сачуван и уписан на било који диск у кластеру. Дакле, било би могуће, али такође и неуобичајено, да се похрани цела једна датотека у кластеру рачунара, а да су сви реплицирани блокови попунили све дискове. Друго, смањује се потреба клијента да комуницира са *Master Node*-ом јер читање и писање на истом блоку захтева само један почетни захтев *Master Node*-а за информације о локацији реплика блокова.

Name Node је средишњи део *HDFS* система датотека. Чува стабло директоријума свих датотека у фајл систему и прати где се преко кластера чувају подаци датотека, не похрањујући саме податке о тим датотекама. Клијентске апликације комуницирају са *Name Node*-ом када год желе лоцирати датотеку или је копирати, преместити или избрисати. *Name Node* одговара на успешне захтеве враћајући листу релевантних *Data*

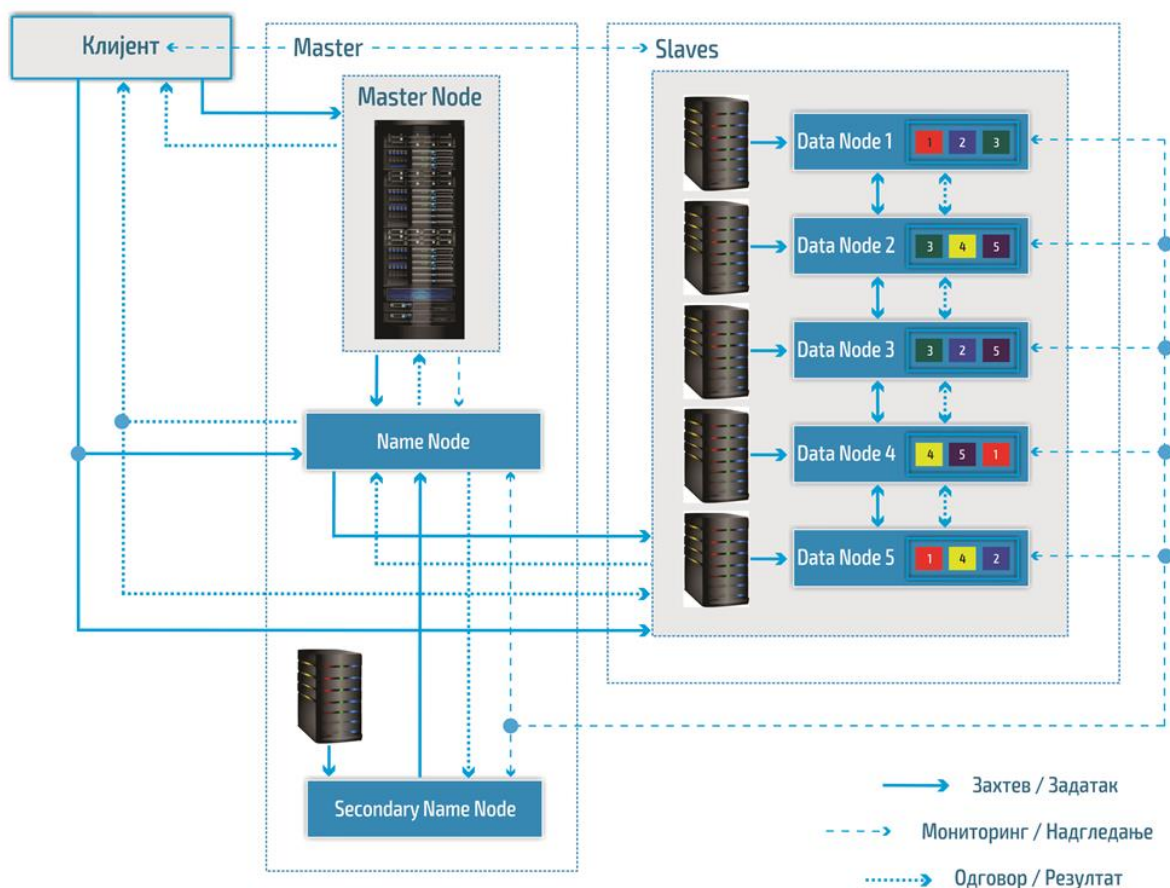
Node сервера у којима су подаци похрањени и одређује мапирање блокова на *Data Node*-овима (Usama et al., 2017). Обзиром да је *Name Node* јединствена тачка могућих кварова за *HDFS* кластер; у ком случају би фајл систем био искључен, *Secondary Name Node* извршава контролне операције и мониторинг контролних тачака. Чува последњу контролну тачку у директоријуму која је структурирана на исти начин као и директоријуму примарног *Name Node*-а. Покреће се на другој машини у односу на примарни *Name Node* јер су меморијске потребе истог редоследа као и примарног *Name Node*-а. *Secondary Name Node* уређује датотеке дневника задржавајући величину уређивања унутар директива ограничења. Наиме, почетак мониторинга и надгледања контролне тачке контролишу два кључна конфигурациона параметра:

- `dfs.namenode.checkpoint.period` – време које одређује максимално кашњење између две узастопне контролне тачке, које је подразумевано постављено на 1 сат, и
- `dfs.namenode.checkpoint.txns` – број неозначених трансакција на *Name Node*-у, који је подразумевано постављен на 1 милион, што ће приморати хитну контролну тачку, чак и ако период контролне тачке није постигнут.

Data Node – чува податке на *HDFS* слоју система. У пројектовању хардверског система паметне библиотеке, функционалан фајл систем има више од једног *Data Node*-а са дистрибуираним репликама података копираним преко њих, притом обезбеђујући толерантност на грешке где неисправност рада једног *Data Node*-а неће уништити датотеку нити успорити или зауставити рад система. Приликом покретања система, *Data Node* се повезује са *Name Node*-ом и прима директиве које га упућују на:

- реплицирање блокова датотеке на друге нодове у кластеру;
- међусобну комуникацију *Data Node*-ова приликом копирања података;
- извршавање захтева читања / писања од стране клијента;
- чување метаподатака блокова и њихово складиштење у локалном фајл систему;
- слање извештаја *Name Node*-у о статусу свих постојећих блокова;
- дељење података и метаподатака клијенту.

Клијентске апликације могу комуницирати директно са *Data Node*-ом, након што *Name Node* наведе локацију блокова података. Клијент пре експортовања тражених одговора у апликацију, *Master Node*-у шаље захтев за индексе реплике блокова датотеке. *Master Node* даје одговор одговарајућим блоком и локацијом реплика. Клијент бележи / кешира добијену информацију користећи име датотеке и индекс блока као његовог кључа, обезбеђујући непотребне будуће интеракције између клијента и *Master Node*-а.

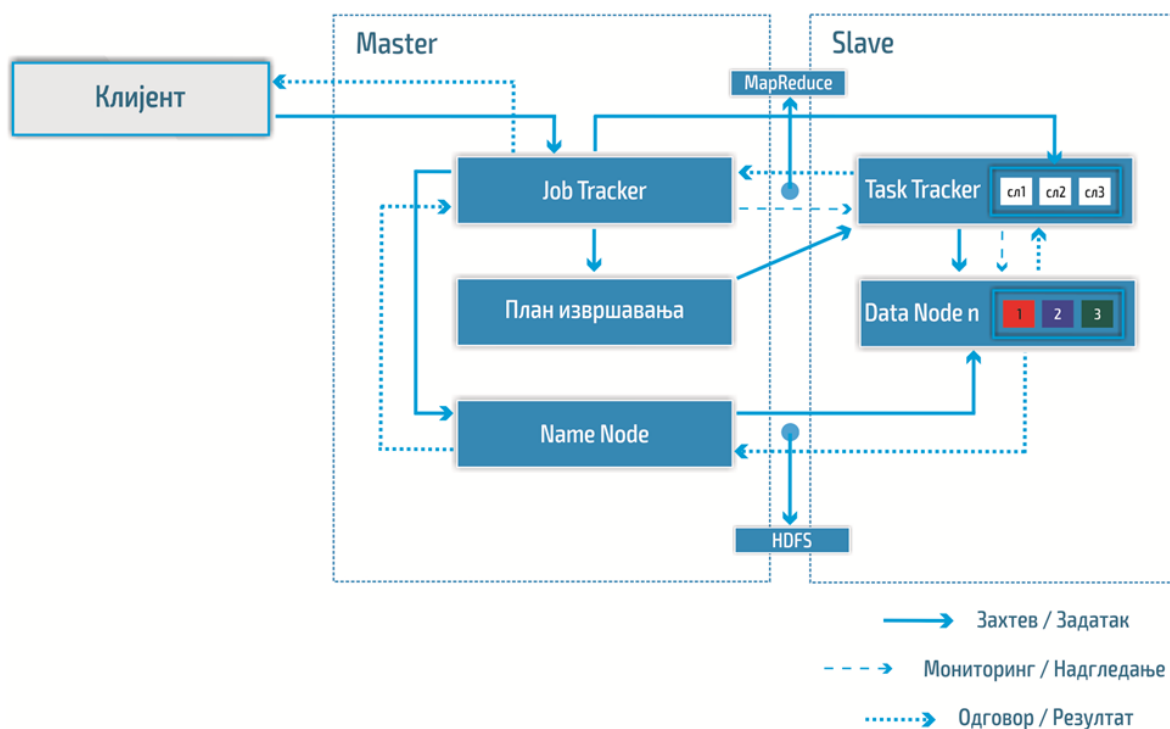


Слика 32. Детаљан приказ пројектоване Master / Slave архитектуре мултинод кластера

Слично томе, *MapReduce* операције које су извршене на инстанци *Task Tracker*-а комуницирају директно са *Data Node*-ом за приступ датотекама. Примера ради, *Task Tracker* се може распоредити на истим серверима на којима се налазе инстанце *Data Node*-а, тако да се *MapReduce* операције изводе у близини података. Сваки *Task Tracker* је конфигурисан са сетом празних слотова (сл1, сл2, сл3, ..., слn), што указује на број задатака које може прихватити, након чега *Job Tracker* тражи празан слот на истом серверу на којем се налази *Data Node* који садржи податке. *Task Tracker* прати покренуте процесе, преузимајући излазне кодове. Када се процес заврши, успешно или не, *Task Tracker* обавештава *Job Tracker*. *Task Tracker* шаље поруке о исправности свог рада, обично сваких неколико минута, *Job Tracker*-у. Ове поруке обавештавају *Job Tracker* о броју доступних слотова, тако да је *Job Tracker* континуирано у току са датим задацима у кластеру и планом извршавања послова који су делегирани.

Job Tracker у екосистему комуницира са *Name Node*-ом како би одредио локацију блокова података. Одговоран је за избор сервера који покрећу обраду података, као и за планирање задатака на дељеном кластеру. Након што клијентска апликација достави задатак *Job Tracker*-у, *Job Tracker* лоцира нодове *Task Tracker*-а са расположивим слотовима и задаје послове изабраном ноду. Исправност рада *Task Tracker* нодова се прате, и ако један од сервера који је укључен у *MapReduce* обраду ради споро у односу

на друге који су обраду завршили, *Job Tracker* покреће другу инстанцу тог истог задатка на другом серверу у кластеру тако да преоптерећени чворови у кластеру неће зауставити нити успорити *MapReduce* посао (Olson, 2010; White, 2012, Wang et al., 2013). Када је посао завршен, *Job Tracker* ажурира свој статус и дистрибуира информације клијентској апликацији (приказано на слици 33).



Слика 33. Управљање операцијама Master / Slave мултинод кластера

4.1.3. СОФТВЕРСКИ СЕРВИСИ ПРОЈЕКТОВАНОГ МОДЕЛА ПАМЕТНЕ БИБЛИОТЕКЕ

За пројектовање и реализацију софтверских сервиса модела паметне библиотеке, изабрана је платформа *Apache Hadoop Hortonworks*. *Apache Hadoop* је широко прихваћена и једна од најповољнијих *big data* софтверских платформи која подржава дистрибуирано процесирање и рачунарску парадигму *MapReduce* која омогућава паралелну обраду велике количине хетерогених података.

MapReduce је програмски модел за обраду и генерисање великих скупова података који се могу прилагодити широком спектру задатака.

Корисници дефинишу програмска израчунавања преко функције мапирања и редуцирања података који аутоматски (Dean and Ghemawat, 2008):

- извршавају паралелну обраду на великим кластерима сервера / рачунара;
- управљају кваровима на машинама; и

- планирају и заказују интер-машинску комуникацију ради ефикасног коришћења мреже, сервера и дискова.

Програмери сматрају да је систем једноставан за употребу. Више од десет хиљада *MapReduce* програма интерно је имплементирано на *Google*-у и просечно се око сто хиљада *MapReduce* послова извршава на *Google*-овим кластерима сваки дан, обрађујући укупно више од двадесет петабајта података дневно (*Dean and Ghemawat, 2008*). *Hadoop* екосистем, са имплементираним *MapReduce* и *HDFS* софтверским сервисима се сматрају најефикаснијим оквиром за управљање великим подацима (*Khan et al., 2017*).

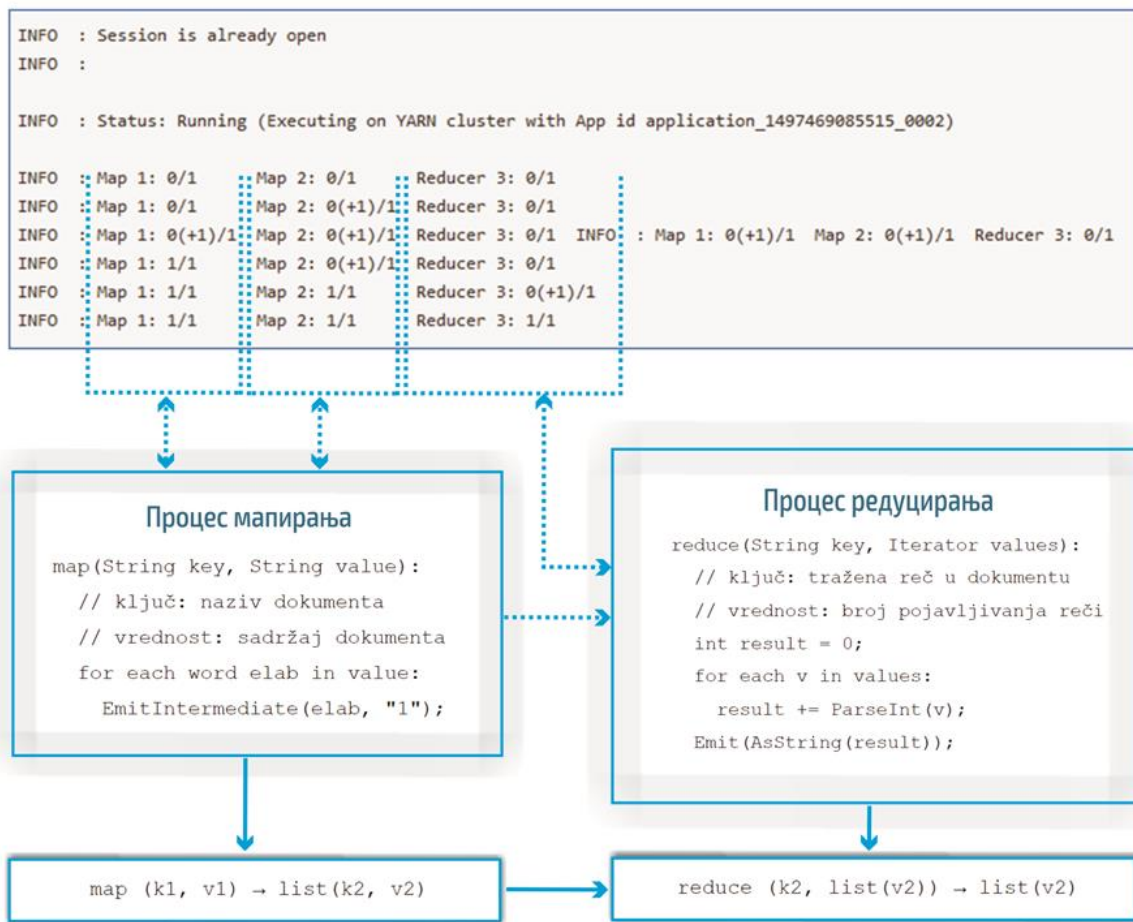
Паралелну обраду вишеструких података, подељених по блоковима у кластеру рачунара извршава *MapReduce* софтверски оквир. *MapReduce* обезбеђује паралелну обраду скупова података које се деле и распоређују у кластеру. *Master Node* додељује задатке *Slave Node*-у и затим прикупља резултате (*Simović, 2018*). Овако дефинисан модел има два главна корака (*Dean and Ghemawat, 2008*) (приказано на слици 34):

- 1) *Map* – мапирање распоређеног посла и
- 2) *Reduce* – редуковање и прикупљање резултата.

Софтверски *MapReduce* сервис пројектованог модела, израчунавање извршава тако што узима скуп улазних парова дефинисаних као кључ / вредност (енгл. *key / value*), и производи скуп излазних парова дефинисаних такође као кључ / вредност. Мапирање узима улазни пар након чега производи скуп посредних (енгл. *Intermediate*) парова; групише њихове вредности које су повезане са истим посредним кључем *K*; и прослеђује их функцији редуковања.

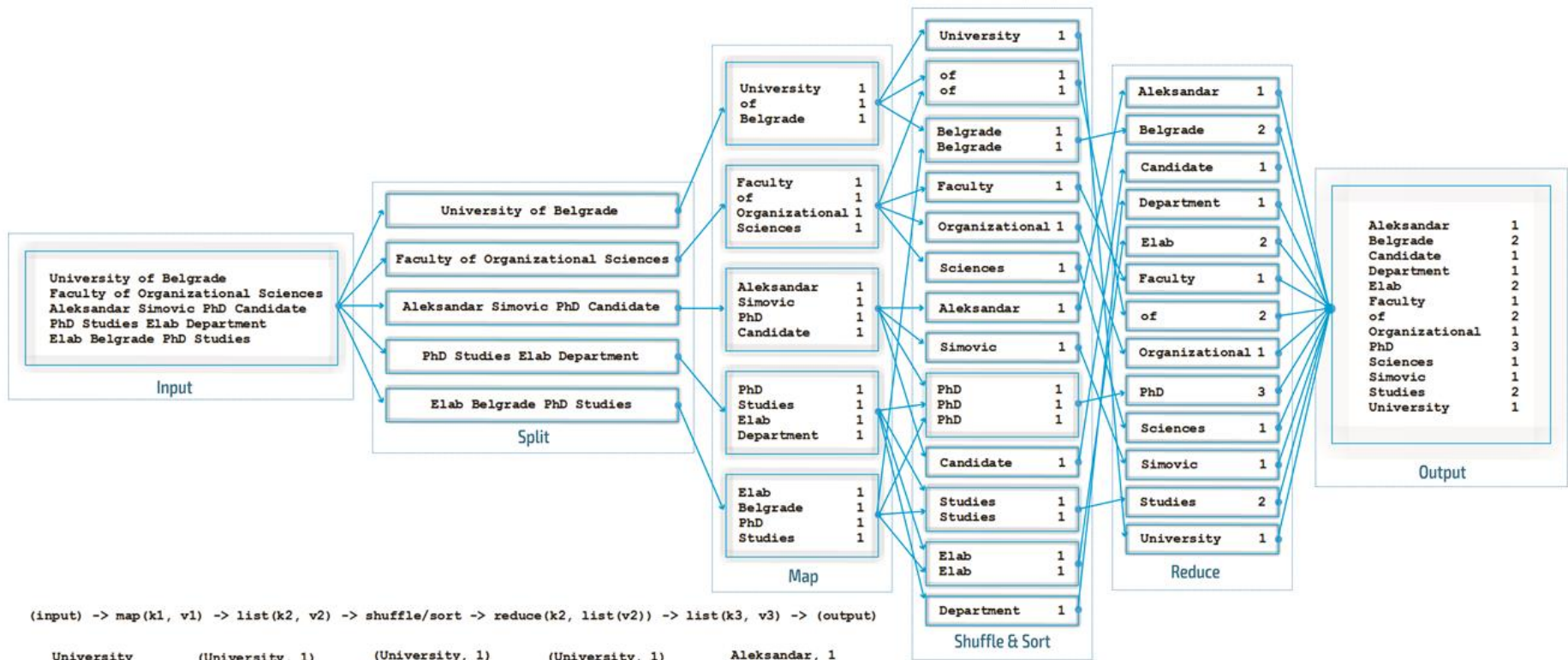
Редуковање са својом функцијом у систему, преузима посредни кључ *K* са укупним скупом вредности тог кључа, након чега их обједињује ради генерисања мањег скупа вредности на излазу. Уобичајено је да ће бити добијена једна или ниједна излазна вредност након функције редуковања.

Посредне вредности се преко итератора (енгл. *Iterator*), који ће за сваки задати посао, поновити поступак израчунавања као *Shuffle & Sort* секвенце дела *MapReduce* софтверског оквира, прослеђују редукиционој функцији. Овако дефинисан *MapReduce* процес омогућава обраду листе вредности које су превелике да би се уклопиле у меморију. Примера ради, уколико разматрамо проблем броја укупног појављивања једне речи или једне фразе у великом скупу података или великој збирци веб докумената, псеудо кôд са *MapReduce* сесијом који решава дат проблем је приказан на слици 34.



Слика 34. Псеудо код у задатој MapReduce сесији (адаптирано према: Dean and Ghemawat, 2008)

Функција мапирања у приказаном примеру емитује сваку реч са придруженим бројем појављивања те речи; док функција редуције, сумира укупан број појављивања емитоване речи. *MapReduce* спецификација објекта се с именима улазних и излазних датотека и опционим параметрима подешавања исписује, након чега се позива функција која се прослеђује објекту спецификације. Наиме, итератор представљен у псеудо коду, у процесу редуцирања, обухвата поступак израчунавања преко *Shuffle / Sort* секвенце *MapReduce* софтверског сервиса пројектованог модела паметне библиотеке. Детаљан приказ секвенци *MapReduce* софтверског оквира као програмабилног интерфејса за приступ вишеструким реплицираним подацима дистрибуираних преко *Data Node*-ова у кластеру је приказан на слици 35.



University	(University, 1)	(University, 1)	(University, 1)	Aleksandar, 1
of	(of, 1)	(of, [1, 1])	(of, 2)	Belgrade, 2
Belgrade	(Belgrade, 1)	(Belgrade, [1, 1])	(Belgrade, 2)	Candidate, 1
Faculty	(Faculty, 1)	(Faculty, 1)	(Faculty, 1)	Department, 1
of	(of, 1)			Elab, 2
Organizational	(Organizational, 1)	(Organizational, 1)	(Organizational, 1)	Faculty, 1
Sciences	(Sciences, 1)	(Sciences, 1)	(Sciences, 1)	of, 2
Aleksandar	(Aleksandar, 1)	(Aleksandar, 1)	(Aleksandar, 1)	Organizational, 1
Simovic	(Simovic, 1)	(Simovic, 1)	(Simovic, 1)	PhD, 3
PhD	(PhD, 1)	(PhD, [1, 1, 1])	(PhD, 3)	Sciences, 1
Candidate	(Candidate, 1)	(Candidate, 1)	(Candidate, 1)	Simovic, 1
PhD	(PhD, 1)			Studies, 2
Studies	(Studies, 1)	(Studies, [1, 1])	(Studies, 2)	University, 1
Elab	(Elab, 1)	(Elab, [1, 1])	(Elab, 2)	
Department	(Department, 1)	(Department, 1)	(Department, 1)	
Elab	(Elab, 1)			
Belgrade	(Belgrade, 1)			
PhD	(PhD, 1)			
Studies	(Studies, 1)			

Слика 35. MapReduce Shuffle & Sort процес

Применом функционалног *MapReduce* програмског оквира, структура података је смештена у меморију чиме су избегнуте честе синхронизације процеса и заузимање ресурса.

Доступност података у паметном библиотечком моделу обезбеђује *HDFS* континуираним надгледањем чворова у кластеру и блоковима којима задужени чворови управљају. Сви блокови подлежу проверама и контролама. Када је блок прочитан, утврђује се исправност (без обзира да ли је снимљена вредност тачна). Ако је блок оштећен, бива замењен једним од његових реплика из другог кластер клона (*Olson, 2010; White, 2012*).

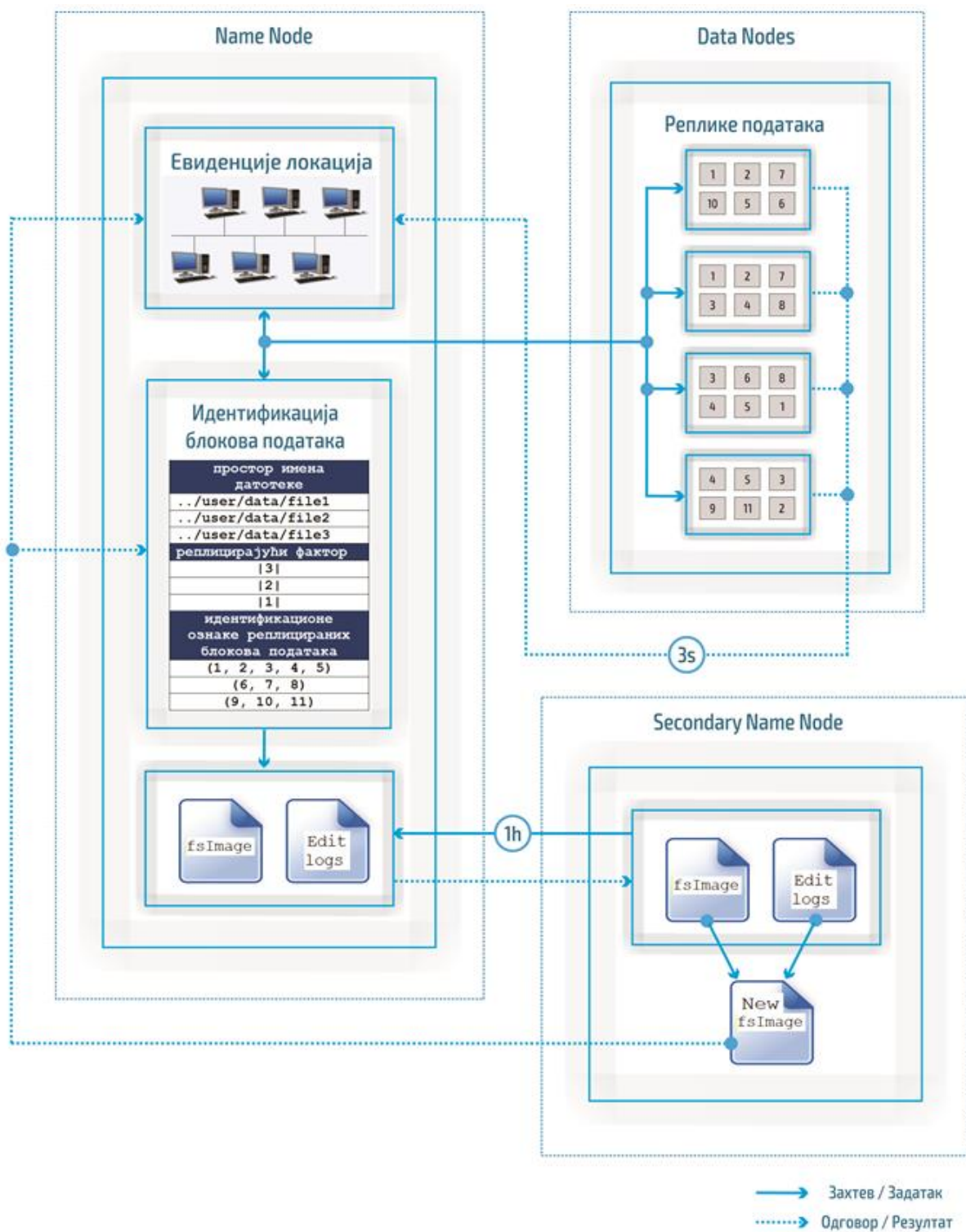
Пројектовани *HDF* систем са могућностима репликације омогућава:

- толеранцију на грешке – где губитак једног чвора неће уништити датотеку; и
- перформансе – где се било који блок може читати са једног или више чворова уз побољшање протока података.

Подаци се чувају преко *HDFS*-а који омогућава скалабилно складиштење и толерантност на грешке. *HDF* систем открива и надокнађује грешке у кластеру раздвајајући долазне датотеке у блокове, чувајући их преко кластера рачунара. Датотеке су подељене у блокове и сваки блок се копира на више *Data Node*-ова. *HDFS* управља мета подацима фајл система и мапирањем података целог кластера које чува у меморији *Name Node*-а као *fsImage* фајл, чија копија се као подршка, по правилу, на сваки сат чува и ажурира на *Secondary Name Node*-у.

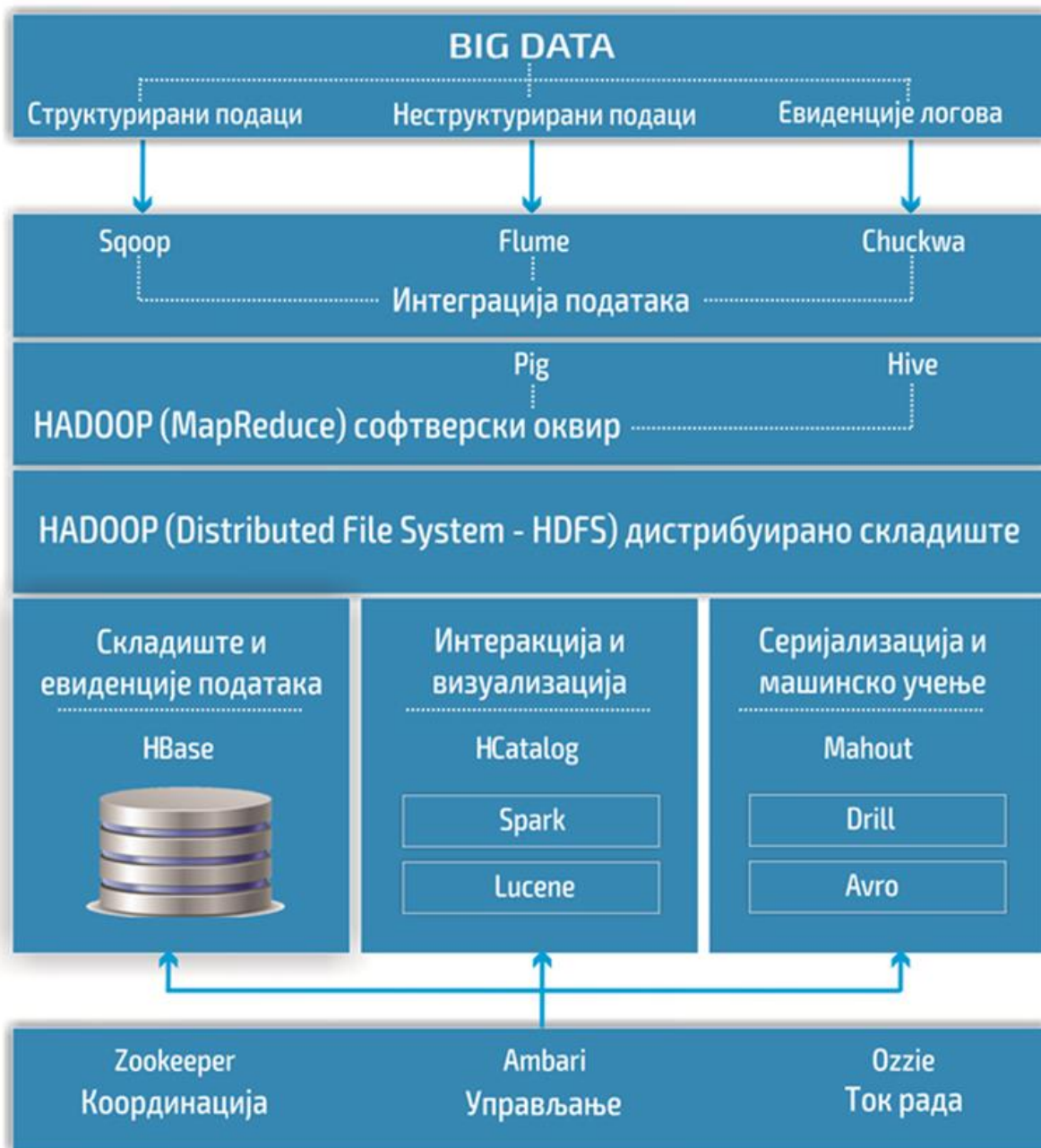
Карактеристична особина *HDF* система је евидентирање позиција свих података у фајл систему путем (приказано на слици 36):

- представе простора имена датотеке или више датотека;
- праћења свих дистрибуираних реплика података улазних датотека ускладиштених преко *Data Node* сервера у кластеру;
- мапирања идентификационих ознака реплицираних блокова података;
- обезбеђивања довољног броја укупних реплика дистрибуираних података;
- маркирања приоритета недовољних реплика података;
- испитивања исправности блокова података којима *Data Node* сервери управљају;
- слања извештаја *Name Node* серверу о исправности рада *Data Node* сервера у кластеру на сваке три секунде;
- маркирања и уклањања већег броја реплика дистрибуираних података са *Data Node* сервера.



Слика 36. HDFS процес пројектованог модела паметне библиотеке

Big data технолошки оквир са модулима и сервисима од значаја за пројектовање big data решења паметне библиотеке је приказан на слици 37.



Слика 37. Big data технолошки оквир модела паметне библиотеке

Hadoop систем садржи различите модуле који обезбеђују различите перформансе и омогућавају администраторима, програмерима и аналитичарима прилагођен и ефикасан рад. Сваки модул има своју специфичну функцију, а више модула се може груписати у посебне категорије. Типично груписање модула које је примењено у пројектовању модела паметне библиотеке је:

- модул за обраду и приступ подацима;
- за увоз података у *Hadoop* фајл систем;
- за складиштење података;
- за интеграцију података;

- за процесирање, анализу и визуализацију података;
- модул за мониторинг и администрацију података; и
- обраду података у реалном времену (*Simović and Ćirović, 2016*).

У пројектованом моделу, систем паметне библиотеке обрађује и електронска издања књига која су доступна студентима за преузимање на основу изабраних предмета током студирања. Систем извршава анализу кључних речи у е-публикацијама и генерише текстуалну датотеку свих кључних речи за дату књигу, након чега се иста прослеђује дистрибуираном складишту *HDFS*-у. Тада систем препоруке пре генерисања листе узима у разматрање како карактеристике садржаја на основу изабране е-књиге, тако и карактеристике свих е-публикација у систему.

Почетни улаз је конвертовање књига у *txt* екстензију. *Pdf* датотеке се пребацују у текстуалне датотеке. Конвертовање једне књиге није захтевно, међутим уколико је потребно конвертовати велики скуп података (књига у е-издању) потребан је систем који то може извести у смањеном и одговарајућем временском периоду.

Псеудо код за конвертовање *pdf* датотеке у *txt* датотеку је следећи (*Simović and Ćirović, 2016*):

Protokol:
konvertovanje pdf-a u txt datoteku

```
Create a pdf
pNum: number of pages in the document
for page=1 to pNum
    text:pdfTextExtractor.getTextFromPage(instance_of_pdfReader, page);
    write the text to the text file
end for
```

Следећи корак је генерисање укупног броја кључних речи дефинисан за сваку електронску књигу у великом скупу података. *MapReduce* псеудо код за генерисање датотеке укупног броја речи у е-публикацији је приказан у наставку:

Protokol:

Map

```
For each book in bookset
{
    T=tokenize(book);
    For each token in T
    {
        WordCount[token]++;
    }
}
sendToReduceStep(WordCount);
```

Protokol:**Reduce**

```
For each WordCount received from MapStep
{
    Add(totalWordCount, WordCount);
}
```

Генерисана текстуална датотека укупног броја речи у електронској публикацији користи се за издвајање свих кључних речи за дату књигу. Да би се то урадило, праг кључ / вредност $\langle key, value \rangle$ пар се узима и вредности се филтрирају и издвајају се кључне речи из генерисане датотеке. Псеудо кôд је следећи:

Protokol:**Generisanje kljucnih reci - WordCount Hadoop**

```
Set String: WordCount file generated from Hadoop
if(string is not Null)
    String1=string split ("\t");
end if
if(string1[0] is not Null)
    if(string1[1]>threshold_value)
        map.put(string1[0], Double.parseDouble(string1[1]));
    end if
end if
for (Map.Entry<String, Double>entry : map.entrySet())
    write(getKey());
end for
```

MapReduce програмирање у библиотечком екосистему врши *Hive* сервис који омогућава задавање и извршавање упита преко великих скупова података и пружа механизам за пројекцију структуре података. Истакнута карактеристика овог сервиса је структура која је усмерена ка аутоматској паралелизацији која омогућава управљање великим подацима у екосистему (*Thusoo et al., 2009; George, 2011*).

Hive сервис се огледа следећим карактеристикама:

- олакшава читање, писање и управљање великим скуповима података који се налазе у дистрибуираном складишту;
- структура се може пројектовати над подацима који су ускладиштени;
- алат командне линије и управљачки програм *JDBC* (енгл. *Java Database Connectivity*) доступни су за повезивање корисника;
- преводи *SQL* упите у *MapReduce* задатке и извршава их на кластеру;
- разлаже *SQL* упите у мапере и редукторе и проналази начин њиховог спајања који ће бити извршен на кластеру;

- примењује синтаксу сличну *SQL*-у – *HiveQL*;
- скалабилан је, продуктиван, оптимизован и интерактиван – може задавати упите преко целог кластера рачунара у односу на традиционалне *RDBM* системе;
- може применити *OLAP* (енгл. *Online Analytical Processing*) орјентисане упите који су лакши за писање у односу на *MapReduce* упите у Јави;
- није прикладан за *OLTP* (енгл. *Online Transactional Processing*) упите;
- генерише релације из обичних текстуалних датотека (нема одвојених табела, страних кључева и сл.);
- генерише мапуре и редукторе тако да нема трансакција, *record-level* ажурирања, убацивања, брисања и сл;
- прилично је сличан са неким *MySQL* екстензијама, као што су прикази (енгл. *views*), где је могуће сместити резултате упита у исте, што омогућава да их следећи упити користе као табелу. Међутим, не као што су материјализовани прикази у базама података јер се у њима складиште копије, *Hive* сервис је логички конструктор приказа јер неће складиштити копију датих податка;
- омогућава спецификацију начина како се структурирани подаци чувају, деле и партиционишу;
- *Shema On Read* разликује *Hive* сервис од релационих база података. У релационим системима се дефинише шема пре него што се податак учита – *Shema On Write*. *Hive* сервис ради потпуно другачије. Узима неструктуриране податке и примењује шему над њима. Подаци су учитани и ускладиштени без икакве структуре и информација које ти подаци носе:

```
CREATE TABLE ratings (
    userID INT,
    bookID INT,
    rating INT,
    time INT)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY '\t'
STORED AS TEXTFILE;
LOAD DATA LOCAL INPATH '${env:HOME}/putanja_do_fajla'
OVERWRITE INTO TABLE ratings;
```

- `LOAD DATA` – начин измештања податка из дистрибуираног фајл система у *Hive* сервис;
- `LOAD DATA LOCAL` – начин копирања податка из локалног фајл система у *Hive* сервис;
- управљање подацима у односу на примену екстерних табела:

```

CREATE EXTERNAL TABLE IF NOT EXISTS ratings (
    userID INT,
    bookID INT,
    rating INT,
    time INT)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'
LOCATION 'putanja_do_fajla';

```

- *Hive* сервис одржава *metastore* и дефинише структуру над неструктурираним подацима који се чувају у *HDFS*-у;
- структура и шема се генерише када се податак прочита;
- *Hive* сервис се примењује преко:
 - *Ambari* сервиса (приказано на сликама у прилогу од A1-A2);
 - интерфејса командне линије (енгл. *Command line interface – CLI*);
 - *Hue* сервиса;
 - *Apache Thrift* сервиса без употребе *OLTP* (енгл. *Online Transactional Processing*) упита; и
 - *Oozie* сервиса.

Примењени софтверски сервиси пројектованог модела паметне библиотеке приказани у *big data* технолошком оквиру на слици 37 су детаљно описани у поглављу 3.1.4. *Big data* сервиси.

Провера успешног покретања и конфигурације примењених сервиса је приказана на сликама у прилогу од A3-A10.

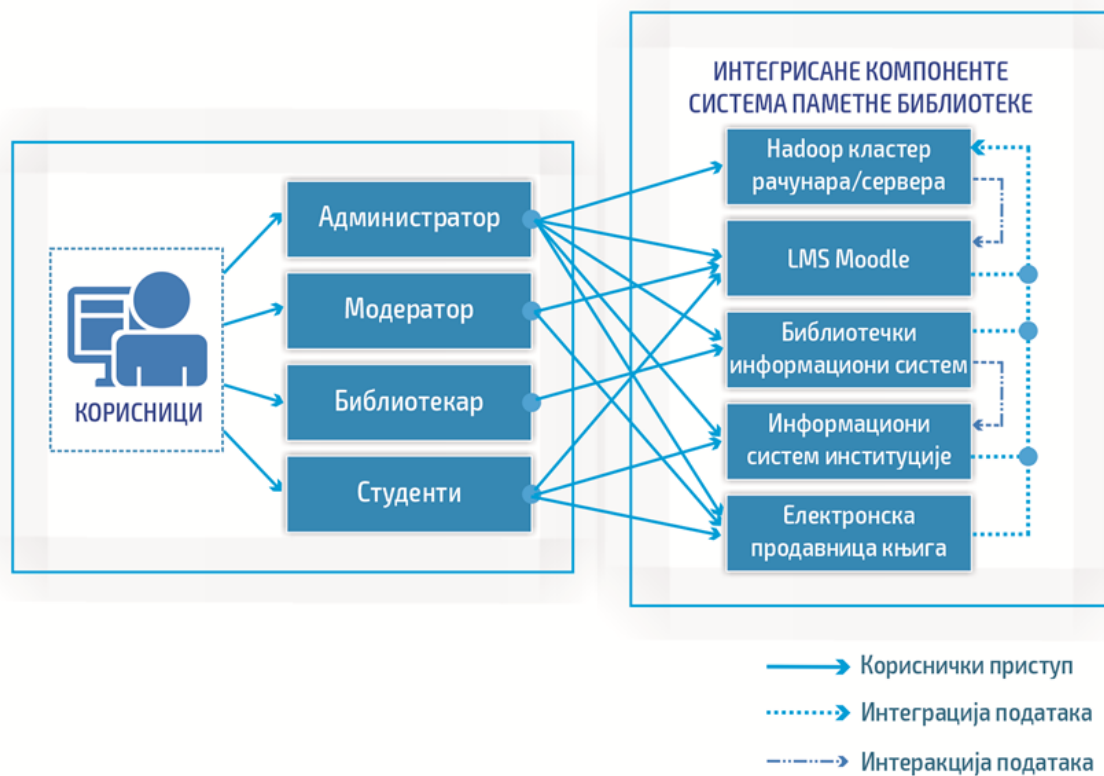
4.1.4. АПЛИКАЦИЈЕ СИСТЕМА ПАМЕТНЕ БИБЛИОТЕКЕ

Апликацијама система паметне библиотеке имају приступ корисници (приказано на слици 38) са дефинисаним нивоом приступа интегрисаним компонентама и могу се поделити у следеће групе: (1) администратор; (2) модератор; (3) библиотекар; и (4) студенти.



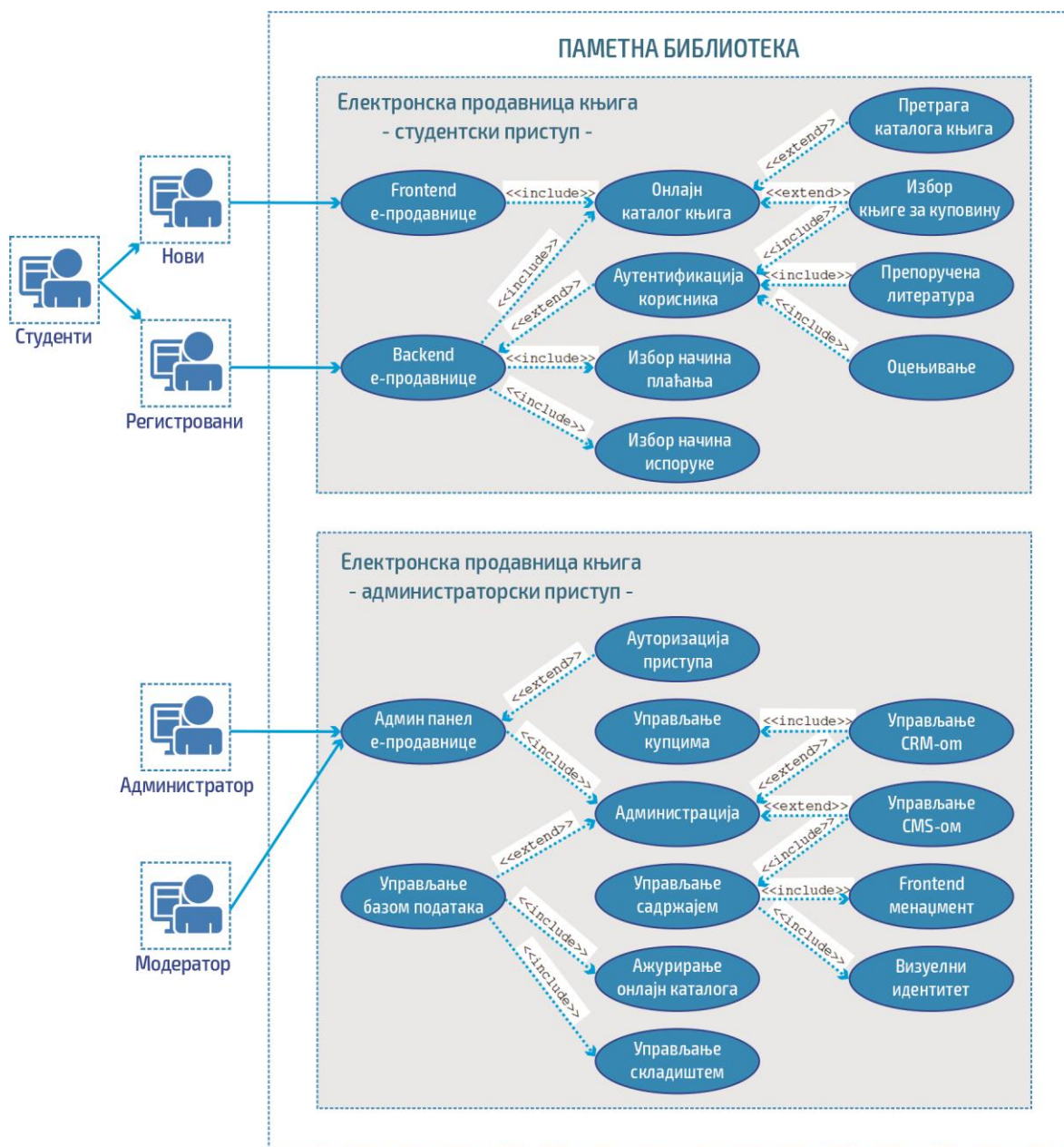
Слика 38. Кориснички приступ систему паметне библиотеке

Детаљан приказ приступа корисника апликацијама компонената система са посебно дефинисаним нивоом приступа интегрисаним компонентама система паметне библиотеке и међусобним релацијама интеграције и интеракције података је приказан на слици 39.

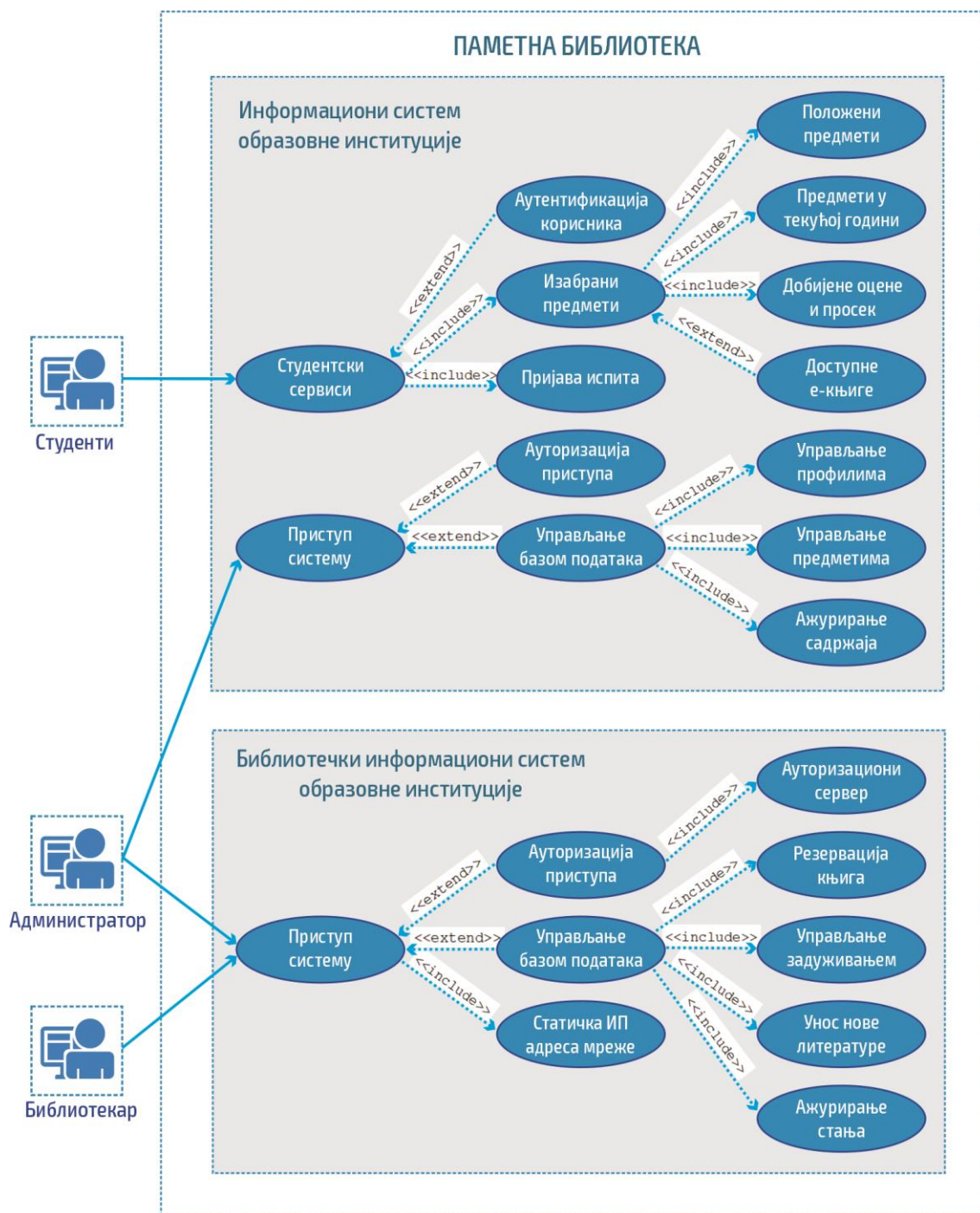


Слика 39. Приступ корисника интегрисаним компонентама система паметне библиотеке са међусобним релацијама интеграције и интеракције података

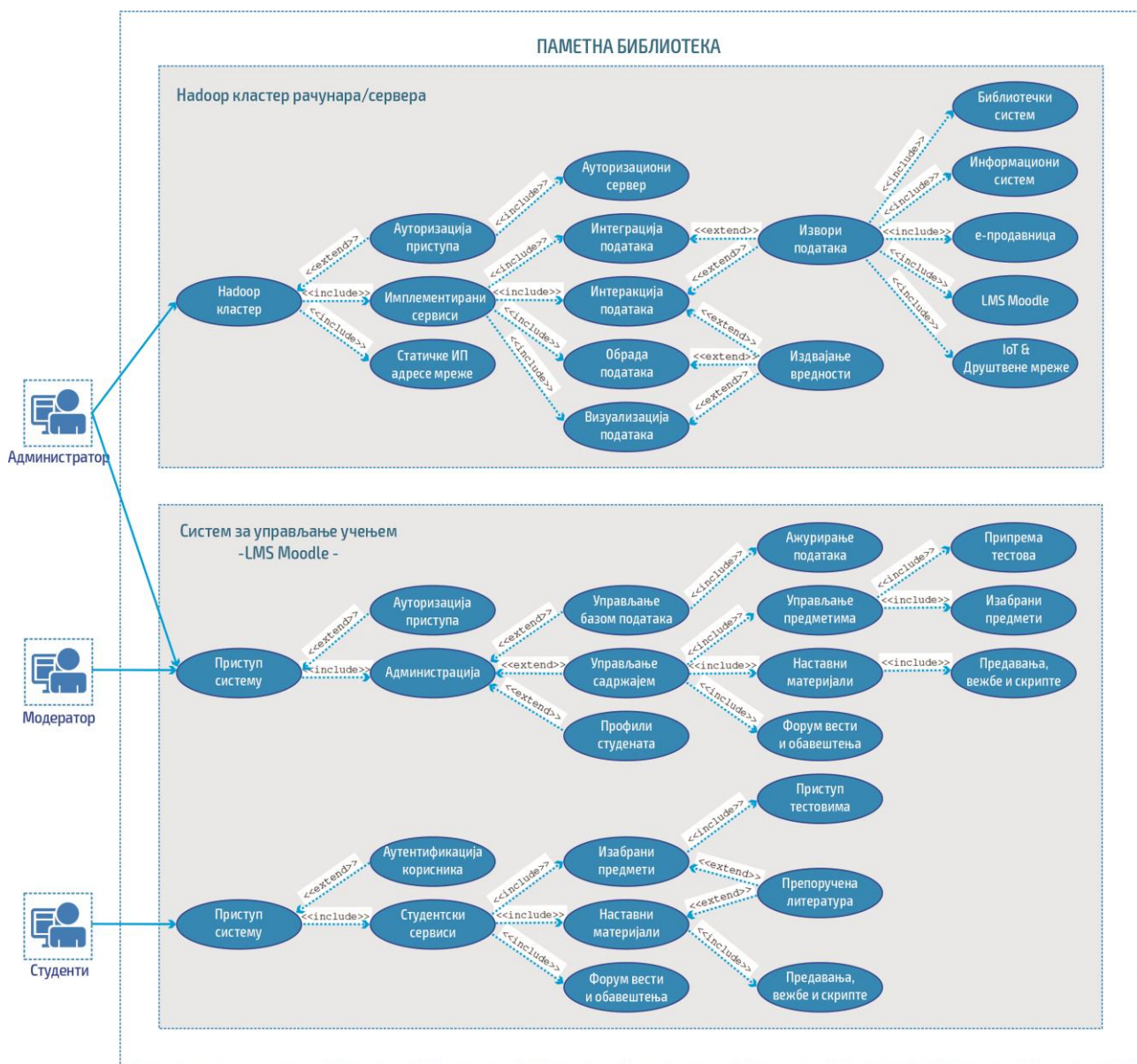
Кориснички дијаграми (енгл. *Use Case*) приступа апликацијама система и интегрисаним компонентама система паметне библиотеке су приказани на сликама 40, 41 и 42.



Слика 40. Кориснички дијаграм (енгл. *Use Case*) приступа апликација система паметне библиотеке – Електронска продавница књига



Слика 41. Кориснички дијаграм (енгл. Use Case) приступа апликација система паметне библиотеке – Информациони и библиотечки информациони систем образовне институције



Слика 42. Кориснички дијаграм (енгл. Use Case) приступа апликација система паметне библиотеке – Hadoop кластер & LMS Moodle

Апликације система паметне библиотеке укључују интеграцију скупова података у централни део екосистема кластера рачунара. Ови различити скупови података носе информације о:

- задуженим књигама студената из библиотеке образовне институције;
- књигама у електронској форми које су расположиве студентима за преузимање преко информационог система образовне институције;
- евиденцијама сервера електронске продавнице књига образовне институције које садрже информације о преференцијама студената и других корисника;
- изабраним курсевима студената у текућој школској години путем *LMS Moodle* платформе.

Са сваком новим приступом регистрованог корисника (студента) на *LMS Moodle* платформу,

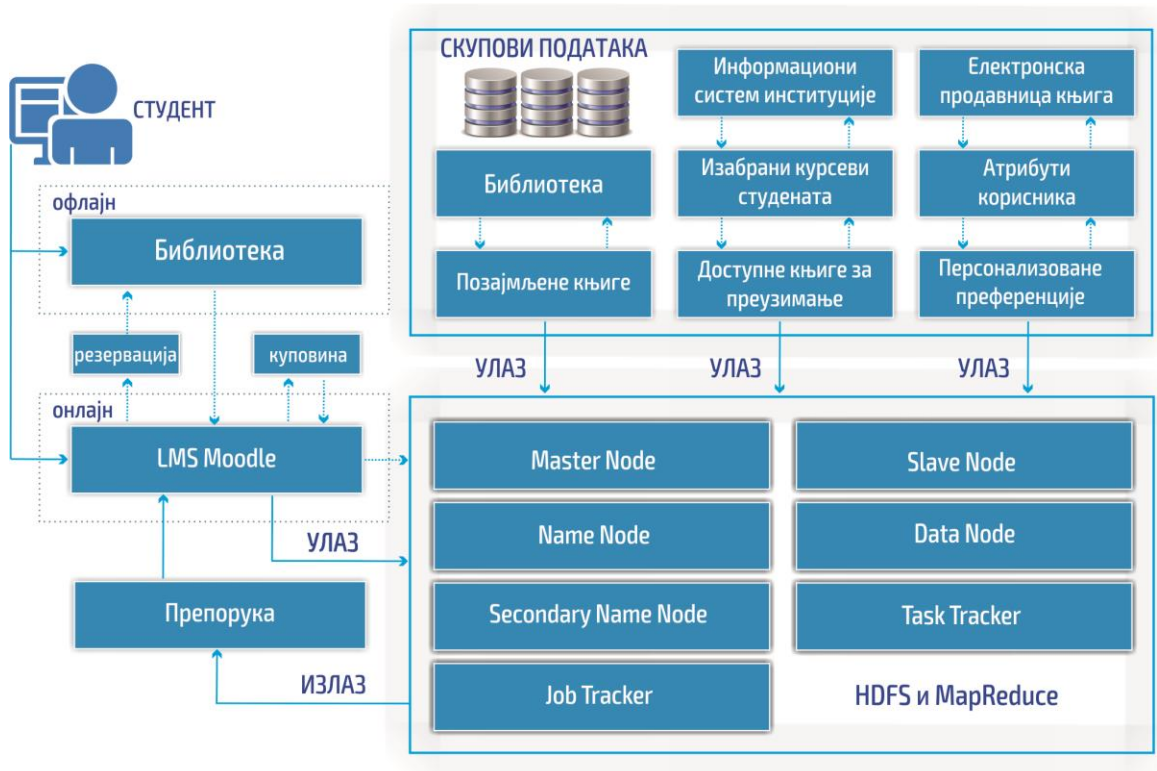
Hadoop екосистем анализира следеће:

- податке о атрибутима књига из евиденције задужења из библиотеке образовне институције;
- податке из информационог система образовне институције у вези са изабраним курсевима студената у текућој школској години;
- историје куповина и атрибуте корисника из евиденција сервера електронске продавнице књига образовне институције.

Након анализе података, *Hadoop* екосистем извршава колаборативно филтрирање (описано у поглављу 3.2.2. Системи колаборативног филтрирања) над интегрисаним скупом података након чега студент добија препоруку садржаја од већег интереса, са два расположива решења:

1. Избор 1. Да резервише препоручене књиге у библиотеци образовне институције. У том случају библиотекар извршава резервацију (бележи идентификациони број студента) за одређени временски период за задуживање; и
2. Избор 2. Да купи препоручене књиге преко електронске продавнице књига образовне институције.

На слици 43 је приказан дијаграм тока модела *big data* екосистема паметне библиотеке.



Слика 43. Дијаграм тока модела *big data* екосистема паметне библиотеке

4.2. ИНТЕГРАЦИЈА ПОДАТАКА У ЕКОСИСТЕМ

4.2.1. ПОДАЦИ ИЗ ИНФОРМАЦИОНОГ СИСТЕМА ОБРАЗОВНЕ ИНСТИТУЦИЈЕ

Интегрисан скуп података из информационог система образовне институције за три године студија садржи 160 курсева, седам студијских програма и ~2250 студената, са укупно преко 10 милиона записа. Сваки студент са свог налога за онлајн приступ студентском сервису може преузети од 5 до 15 књига у електронском формату у зависности од одабраних предмета и доступне литературе.

4.2.2. БИБЛИОТЕЧКИ СКУП ПОДАТАКА

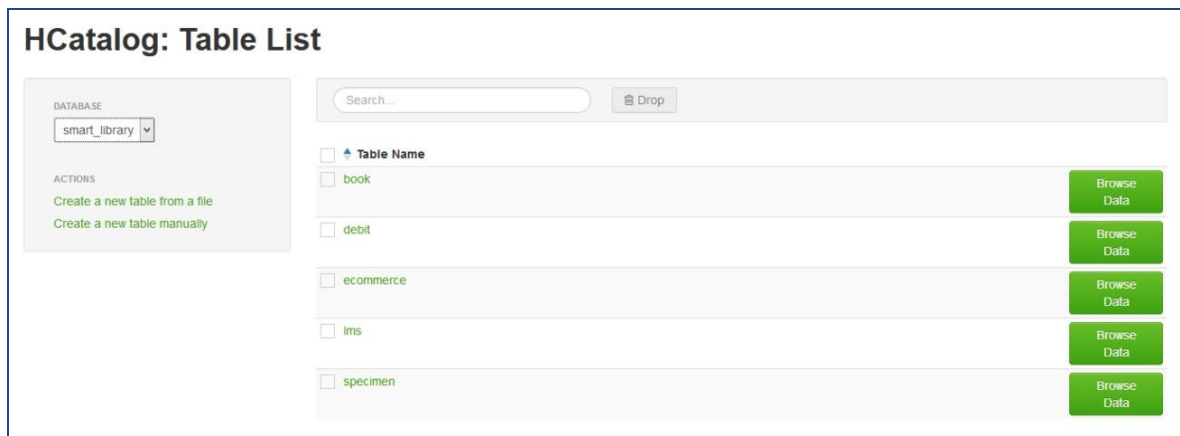
Библиотечки скуп података садржи 470571 корисника и 3955 наслова књига. Евиденција сервера електронске продавнице књига образовне установе која је коришћена садржи ~3900 евалуација, ~450 корисника и ~1495 ставки које су узете за обраду. У анализи интегрисаних скупова података ради њихове обраде, важан фактор је покривеност (*Sarvar et. al.*, 2000). Покривеност је дефинисана као укупан проценат доступних ставки (производа е-каталога) за које ће систем генерисати препоруке.

4.2.3. ПОДАЦИ СА *LMS MOODLE* ПЛАТФОРМЕ

Сваки студент се пријављује на курсеве које је изабрао у текућој школској години путем *LMS Moodle* платформе, што може бити између 10 и 12 курсева. *LMS Moodle* платформа садржи следеће информације: три године основних студија (7 студијских програма), једну годину специјалистичких студија (6 студијских програма) и две године мастер студија (3 студијска програма) – што је укупно 440 курсева, ~3750 студената, ~900 наставника и преко 15 стотина записа дневно.

4.2.4. ЛИСТА ТАБЕЛА УЧИТАНИХ СКУПОВА ПОДАТАКА

Наведени скупови података су учитани у *HCatalog* и спремни су за даљу обраду на слоју управљања екосистемом (приказано на слици 44). Подаци се касније дистрибуирају и преносе у *HBase* модул – сервис који омогућава претраживање, преузимање и анализу података.



Слика 44. HCatalog – листа учитаних табела

4.2.5. ИНТЕГРАЦИЈА ПОДАТАКА ДРУШТВЕНИХ МЕДИЈА

Метод описан у овом поглављу се односи на прикупљање и агрегацију података са сајтова за друштвено умрежавање и обезбеђивање алата за испитивање и анализу тих информација. Агрегација података друштвених мрежа треба да омогући анализу структурираних и неструктурираних података из различитих канала друштвених медија (*Gandomi and Haider, 2015*). Друштвени медији су широк појам који обухвата онлајн платформе које корисницима омогућавају креирање и размену садржаја. Друштвени медији се могу сврстати у следеће типове: друштвене мреже (нпр. *Facebook* и *LinkedIn*); блогови (нпр. *Blogger* и *WordPress*); микроблогови (нпр. *Twitter* и *Tumblr*); друштвене вести (нпр. *Digg* и *Reddit*); дељење видео и фото садржаја (нпр. *Instagram* и *YouTube*); вики платформе (нпр. *Wikipedia* и *Wikihow*); сајтови за питања и одговоре (нпр. *Yahoo! Answers* и *Ask.com*); и странице за преглед садржаја (нпр. *Yelp* и *TripAdvisor*) (*Barbier and Liu, 2011; Gundecha and Liu, 2012; Gandomi and Haider, 2015*).

На слици 45 је приказан метод за агрегацију података и аналитику текста прикупљених са друштвених медија. Аналитика текста (енгл. *Text mining*) се односи на технике које извлаче информације из текстуалних података. Постови са друштвених мрежа, е-поруке, блогови, онлајн форуми, одговори на постављене анкете, евиденције позивних центара, све су примери текстуалних података које компаније и организације могу да складиште и да са њима манипулишу. Аналитика текста укључује статистичку анализу, рачунску лингвистику и машинско учење, и омогућава компанијама да конвертују велике количине људског генерисаног текста у смислене резимее, који подржавају доношење одлука заснованих на доказима. На пример, аналитика текста се може користити за предвиђање цена акција на берзи на основу информација које су анализиране из финансијских вести (*Chung, 2014; Gandomi and Haider, 2015*).

Као што је приказано на слици 45, први аналитички процес који се извршава на СДМ-у (скр. Сервер Друштвених Медија), сакупља податке са друштвених медија из различитих извора. Извори могу укључивати друштвене медије као што су *Facebook* или *LinkedIn*, или микроблог *Twitter*. Процес може филтрирати прикупљене податке по кључној речи, кључној фрази или корисничком идентификационом броју како би се смањио обим укључених извора

друштвених медија (Feng et al., 2016). Процес би могао да филтрира постове на основу одређених тема, на пример „big data“, „електронско пословање“, „електронско образовање“, и сл.

Модул за прикупљање података у оквиру аналитичког процеса СДМ-а извршава обраду информација која укључује анализу друштвених медија како би се извукли ентитети као што су веб локације, имена корисника, ID корисника, ознаке и карактеристике тема и постова. У модулу за анализу сентимента детектује се задовољство корисника у друштвеним медијима тако што се препознају позитивне и негативне речи и ставови корисника на основу изабране теме. Корелација између задовољства корисника и кључне речи може да варира у зависности од извора друштвених медија.



Слика 45. Метод за агрегацију података и аналитику текста прикупљених са друштвених медија

Процес који се извршава у модулу за анализу сентимента, прикупља и архивира поруче са друштвених медија које изражавају мишљење корисника. У следећој инстанци извршавања задатака на аналитичком серверу друштвених медија, процес који се одвија у модулу за категоризацију, анализира друштвене медије како би категорисао медиј по једној или више тема. На пример, процес може укључити наслов књиге, име аутора, оцену квалитета према

субјективном задовољству и мишљењу корисника (нпр. „добар“, „лош“ или „нејасан“) и сл. Такве теме могу бити унапред дефинисане, или процес може динамички одредити теме консолидовањем порука друштвених медија од више корисника. Процес може да одреди одређене теме према приоритету или важности. На пример, процес може да додели теми као што је „одлично написана књига“ виши приоритет од „нејасно написана књига“.

Следећи задатак аналитике СДМ-а је категоризација корисника. Процес анализира ставове, мишљења и коментаре на друштвеним медијима како би категорисао кориснике повезане са сваким постом према демографским особинама. Категорије могу укључити годиште, пол, годину студија, интересе (нпр. електронско пословање, информационе технологије и сл.) или локацију корисника (нпр. општина, град, регија, држава). Процес може да одреди такве информације из података о профилу корисника или из садржаја постова на друштвеним медијима. Обрада података у овом модулу за категоризацију корисника додатно укључује одређивање утицаја појединачних корисника унутар њихове демографске групе. Подаци са друштвених медија се затим архивирају. Са високом прецизношћу систем архивира корисничке профиле и друштвене везе корисника повезаних са друштвеним медијима.

Након архивирања података, аналитички процес који се извршава на СДМ-у, индексира архивирани друштвене медије помоћу једног или више својстава. Процес индексира податке како би омогућио ефикасно проналажење друштвених медија према својствима која могу укључити, на пример, кључне речи, кључне фразе, задовољства корисника, или демографске податке корисника (*Feng et al., 2016*).

Подаци са друштвених медија, који су на овакав метод прикупљени, анализирани, архивирани и индексирани, могу бити искоришћени у паметној библиотеци и утицати на кориснике на више начина. На пример, ако се из индексираних података извуку информације о кориснику са одређеним идентификационим бројем са позитивним мишљењем о одређеном наслову прочитане књиге која је у корелацији са аутором исте, сродна тема друге књиге истог аутора треба бити обухваћена препоруком садржаја путем система препоруке паметне библиотеке.

4.2.6. ПРИМЕНА *IoT* ТЕХНОЛОГИЈА У МОДЕЛУ ПАМЕТНЕ БИБЛИОТЕКЕ

Паметно управљање процесом задуживања и контроли доступности књига извршава се применом *IoT* технологија. *RFID* и уграђене електронске ознаке на књигама, као бесконтактна и аутоматизована технологија идентификације, у предложеном систему омогућава аутоматско идентификовање објеката (књига) путем сигнала радио фреквенције, обезбеђујући праћење и ажурирање података у реалном времену.

WiFi технологија повезује мобилне уређаје студената са постављеним терминалима на полицама у библиотеци. Високо фреквентни сигнал обезбеђује радијус од ~300 метара и пропусним опсегом од ~37.5Mb/s.

Предложени модел паметне библиотеке примењује *RFID* технологију за препознавање књига и за читање информација са уграђених ознака. *WiFi-RFID* читач прочитане информације са

ознака прослеђује бази података *IoT* серверу путем бежичне мреже, као што се и размена података између мобилних уређаја корисника и сервера обавља, са циљем постизања функционалне претраге књига, управљања залихама и ажурирања ускладиштених података (приказано на слици 46).

У предложеном *IoT* систему сваки *RFID* читач је задужен за одређени опсег књига на полицама библиотеке. Функција претраге књига се обавља путем мобилног уређаја корисника применом пројектоване апликације за мобилне уређаје. Студент или други корисник библиотечких услуга, уноси упит за претрагу као што је: наслов књиге, ИСБН број књиге или аутор/и. Упит се даље преноси путем бежичне мреже *IoT* серверу који обрађује захтев и шаље одговор кориснику да ли је тражена књига доступна за преузимање и задуживање, и уколико јесте, где је њена тачна локација на полици у библиотеци образовне установе. Како би била интуитивнија, мобилна апликација на страни клијента може да генерише мапу путем слике водича ка позицији тражене књиге.



Слика 46. Дијаграм тока преноса података о локацијама штампаних издања књига у библиотеци

Када студент пронађе тражену књигу и жели да је задужи у библиотеци или да је врати током предвиђеног рока за задуживање, књигу приноси *RFID* читачу смештеном на полици ради очитавања информација на основу бесконтактних *RFID* карактеристика и електронских ознака. Високо-фреквентни, пасивни *RFID* читач прослеђује информације које су похрањене у е-ознакама бази података *IoT* сервера, који применом *Storm* модула *Hadoop* екосистема ажурира податке у реалном времену о новом стању залиха и обавља задуживање студента за преузете књиге, или раздужује студента уколико су књиге враћене.

У функцији сигурности, *RFID* систем детекције се поставља на врата улаза / излаза из библиотеке како би проверио следеће:

- да ли се студент пријавио приликом уласка у библиотеку са својом *RFID* картицом и
- да ли су информације са ознака на књигама ажуриране на *IoT* серверу приликом преузимања (задуживања) књига у библиотеци – као потврда да ли је студент преузету књигу прочитао на пасивном *RFID* уређају на полици. Уколико информације нису измењене, *RFID* систем са улаза / излаза, шаље поруку упозорења студенту, као и запосленом у библиотеци, да књиге нису прочитане путем *RFID* терминала на полици библиотеке.

Са друге стране, када се књиге враћају у библиотеку, након пријаве студента на улазу, запослени у библиотеци у систему добија информације о изнајмљеним књигама и временским роковима за задуживање. Ради тачности и истоветности информација, систем паметне библиотеке упоређује податке са *RFID* картице студента, *RFID* ознака на књигама са базом података *IoT* сервера. Систем детекције на улазу је опремљен, како *RFID* терминалом, тао и алармом. Аларм ће се активирати у два случаја:

1. уколико не постоји запис са *RFID* уређаја на полици о изнајмљеним књигама, и
2. уколико се запис не поклапа са записом евиденционе базе података – чиме се ефикасно може избећи губљење књига.

4.3. МЕРЕЊЕ ПЕРФОРМАНСИ МОДЕЛА

Евалуација система је спроведена на два начина. За прву евалуацију, *big data* аналитика је коришћена за утврђивање да ли су најрелевантније књиге (књиге које су највише препоручиване студентима путем *LMS Moodle* платформе) применом система препоруке паметне библиотеке која је заснована на *big data* технологијама чешће изнајмљиване из библиотеке образовне институције у односу на исти период претходне године.

За другу евалуацију система, изабран је упитник као активност за процену модела паметне библиотеке заснованог на *big data* технологијама. Упитник је коришћен да утврди да ли је препорука која је заснована на *Hadoop* екосистему са интегрисаним вишеструким подацима прикупљених из више извора прикладнија за студенте, на основу њихових утисака, у односу на систем препоруке електронске продавнице књига образовне институције.

5. ПРИМЕНА РАЗВИЈЕНОГ МОДЕЛА

5.1. ПЛАН ПРИМЕНЕ РАЗВИЈЕНОГ МОДЕЛА

План примене развијеног модела обухвата *big data* технолошки оквир са модулима и сервисима од значаја за покретање, тестирање и евалуацију *big data* модела паметне библиотеке. Наведени скупови података су прикупљени из четири независна извора и читани у *HCatalog* ради даље анализе и њихове обраде на слоју управљања екосистемом.

У наставку је приказан процес *big data* аналитике над интегрисаним скупом података и извршена је евалуација решења.

5.2. ИМПЛЕМЕНТАЦИЈА РЕШЕЊА

Први упит који је извршен над подацима паметне библиотеке (приказан испод) има за циљ да прикаже кориснике библиотеке који су изнајмили одређене књиге у датом временском периоду 2017. године, и да припреми скуп података за даљу обраду:

```
SELECT debit.person_id,book.title,  
COUNT(debit.person_id) OVER (PARTITION BY debit.person_id ORDER BY debit.person_id  
ROWS BETWEEN UNBOUNDED PRECEDING AND UNBOUNDED FOLLOWING)  
FROM debit JOIN book ON (debit.book_id=book.row)  
WHERE debit.taking_date>'2017-01-01' AND debit.return_date<'2017-06-14';
```

Претраживање, избор и приказ најприкладнијих ставки из укупног садржаја паметне библиотеке су извршени у следећих шест корака (*Simović*, 2018):

Корак 1: Резултат извршеног упита приказује наслове књига и идентификациони број (*ID*) корисника који је задужио највише књига у одређеном временском периоду (приказано у табели 3).

Табела 3. Идентификациони број корисника и наслови књига

<i>ID</i> корисника	Наслови књига изнајмљених из библиотеке
376583	Психолошки услови трансфера учења
376583	Фотонапонска постројења: планирање
376583	Соларне технологије: топлотни и фотоелектрични системи
376583	Физика: механика чврстих

Корак 2: Мониторинг и управљање екосистемом извршава *Ambari* сервис који омогућава инсталирање и праћење *Hadoop* сервиса у кластеру рачунара.

Интегрисан модул извршава (*White, 2012*):

- координацију дистрибуираних апликација;
- синхронизацију и централизацију сервиса у кластеру; и
- мониторинг међусобно независних *Hadoop* послова.

Резултат извршеног израчунавања на основу задатог упита из *Ambari* сервиса је приказан на слици 47.

The screenshot shows the Ambari Hive Query Editor interface. The query editor contains the following SQL query:

```
1 SELECT zaduzenje.id_osobe, knjiga.naziv,  
2 COUNT(zaduzenje.id_osobe) OVER (PARTITION BY zaduzenje.id_osobe ORDER BY zaduzenje.id_osobe  
3 ROWS BETWEEN UNBOUNDED PRECEDING AND UNBOUNDED FOLLOWING)  
4 FROM zaduzenje JOIN knjiga ON (zaduzenje.id_knjige=knjiga.red)  
5 WHERE zaduzenje.datum_uzimanja>'2017-01-01' AND zaduzenje.datum_vracanja<'2017-06-14';
```

The query process results are displayed in a table below the query editor:

zaduzenje.id_osobe	knjiga.naziv	count_window_0
376417	Електричне машине: збирка решених задатака	3
376417	Електричне машине: допуне	3
376417	Електричне машине	3
376558	Termoelektrane	1
376583	Психолошки услови трансфера учења	4
376583	Fotonaponska postrojenja: planiranje	4
376583	Solarne tehnologije: toplotni i fotoelektrični sistemi	4
376583	Физика: механика чврстих	4

Слика 47. Приказ резултата израчунавања у *Ambari* сервису паметне библиотеке

Корак 3: Модел паметне библиотеке повезује идентификациони број студента са његовим налогом електронске поште из следећих екстерних и интерних извора:

- из информационог система образовне институције;
- из библиотеке образовне институције;
- електронске продавнице књига образовне институције; и
- *LMS Moodle* платформе.

Корак 4: Путем онлајн студентског сервиса информационог система образовне институције користећи своје налоге за приступ, студенти могу преузети књиге у електронском формату на основу одабраних предмета за текућу и претходну школску годину. Електронске књиге које су доступне за преузимање у текућој школској години по студенту могу бити у распону од 5 до 15.

Корак 5: Реализован систем препоруке електронске продавнице књига образовне институције примењује колаборативно филтрирање према производима (*Linden et al.*, 2003; *Cantador et al.*, 2008; *Simović*, 2014). При сваком новом пријављивању корисника, систем анализира историју куповине, атрибуте производа и персоналне преференције активног корисника како би пронашао садржај од интереса у електронском каталогу и генерисао листу препорука. Колаборативно филтрирање засновано на производима упоређује сваку куповину корисника, рангира производе према сличности, и њиховим комбиновањем генерише препоруку. Сваком кориснику који је извршио куповину или оценио производе из електронске продавнице књига образовне институције, систем препоручује три књиге.

Корак 6: Приликом нове пријаве корисника на *LMS Moodle* платформу, *big data* систем препоруке паметне библиотеке обрађује податке:

- из библиотеке образовне институције;
- из информационог система образовне институције;
- из евиденција сервера електронске продавнице књига образовне институције; и
- *LMS* система, ради генерисања препоруке.

На слици 48 је приказана генерисана препорука студенту са истим идентификационим бројем коју *Hadoop* екосистем генерише на *LMS Moodle* платформи користећи четири независна извора и четири интегрисана скупа података.

Студент притом добија следећа два избора:

- (1) да резервише препоручене књиге у библиотеци образовне институције, или
- (2) да купи препоручене књиге путем електронске продавнице књига образовне институције.

Алгоритам израчунава сличност ставки ради одређивања најпогоднијих за корисника на основу укупног броја ставки четири независна система. Начин на који алгоритам израчунава сличност ставки са свим осталим унутар *HCatalog*-а је приказан у следећем псеудо коду (адаптирано према: *Linden et al.*, 2003; *Simović*, 2014; и надограђено према: *Simović*, 2018):

```
HCatalog = union ( ProductCatalog1, ProductCatalog2,
                  ProductCatalog3, ProductCatalog4[...])
```

Where the ProductCatalog1 is the library data; the ProductCatalog2 is the IS data of the educational institution; the ProductCatalog3 is the University online bookstore server logs; the ProductCatalog4 is the LMS data;

Where the Selected Item is: Item borrowed from the library; or available for download from the IS of the educational institution; or purchased Item from the University online bookstore; or the student preferences in the current school year;

```
For each Item I1 in HCatalog
  For each User U who Selected Item I1
    For each Item I2 Selected by User U
      Record that a user Selected Item I1 and Item I2
For each Item I2
  Compute the similarity between Item I1 and Item I2
```



Слика 48. Систем препоруке паметне библиотеке заснован на big data технологијама, визуализован на LMS Moodle платформи

Три наслова књига које big data систем препоруке паметне библиотеке генерише посматраном кориснику са истим идентификационим бројем путем LMS Moodle платформе применом big data технологија приказани су у табели 4.

Табела 4. Систем препоруке паметне библиотеке

ID корисника	Препоручени наслови књига применом система препоруке паметне библиотеке применом <i>big data</i> технологија
376583	Увод у проналажење информација на вебу
376583	Одрживи развој
376583	Електронско банкарство

5.3. ЕВАЛУАЦИЈА РЕШЕЊА

У првој евалуацији коришћена је *big data* аналитика да би се утврдило следеће:

- које књиге су најчешће препоручиване студентима приликом евалуације прототипа система; и
- колико пута су препоручене књиге изнајмљене из библиотеке образовне институције у 2016. години у односу на 2017. годину.

Циљ анализе је да се процени потенцијал *Hadoop* екосистема препоруке паметне библиотеке који би допринео побољшању пословног учинка организације, повећању корисничке употребе и поверења у систем.

Следећи упит има за циљ да прикаже које су књиге највише препоручиване студентима путем *LMS Moodle* платформе засноване на *Hadoop* екосистему:

```
SELECT recommender.book_id,book.name,
COUNT(recommender.book_id) OVER (PARTITION BY recommender.person_id ORDER BY
recommender.book_id
ROWS BETWEEN UNBOUNDED PRECEDING AND UNBOUNDED FOLLOWING)
FROM recommender JOIN book ON (recommender.book_id=book.count);
```

Резултати извршеног упита из *Ambari* сервиса су приказани у прилогу Б1 и табели 5.

Табела 5. Књиге које су највише препоручиване студентима путем *LMS Moodle* платформе засноване на *Hadoop* екосистему

ID књига	Наслови књига	Број пута
1016	Основи електронике и телекомуникација	3
844	Хидраулика: увод са примерима управљања	2
1002	Моторна возила 1: општи и теоријски део	2
138	Електричне машине: за трећи разред електротехничке школе	3
1463	Комуторни мотори	3

1497	Термоелектране	3
103	Физика: механика чврстих	6
118	Трансформатори	6
571	Психолошки услови трансфера учења	6
1916	Енергетски трансформатори и генератори	6
2113	Фотонапонска постројења: планирање	6
2275	Соларне технологије: топлотни и фотоелектрични системи	6
7	Основи рачунарске технике	2
264	Математички приручник	2

Следећа два упита су извршена за два одвојена временска периода (за 2016. и 2017. годину) која су затим упоређена.

Први упит одређује колико пута су препоручене књиге изнајмљене из библиотеке образовне институције у датом временском периоду 2016. године, пре него што је прототип модела паметне библиотеке активиран.

```
SELECT recommender.book_id, book.name,
COUNT(recommender.book_id) OVER (PARTITION BY recommender.book_id ORDER BY
recommender.book_id
ROWS BETWEEN UNBOUNDED PRECEDING AND UNBOUNDED FOLLOWING)
FROM recommender JOIN book ON (recommender.book_id=book.count)
WHERE recommender.book_id IN ('1016', '844', '1002', '138', '1463', '1497', '103', '118', '571',
'1916', '2113', '2275', '7', '264')
AND recommender.date_taken>'2016-1-1' AND recommender.date_return<'2017-1-1';
```

Резултати извршеног упита су приказани у прилогу Б2.

Други упит одређује колико пута су препоручене књиге изнајмљене из библиотеке образовне институције у датом временском периоду 2017. године, током тестирања прототипа модела паметне библиотеке.

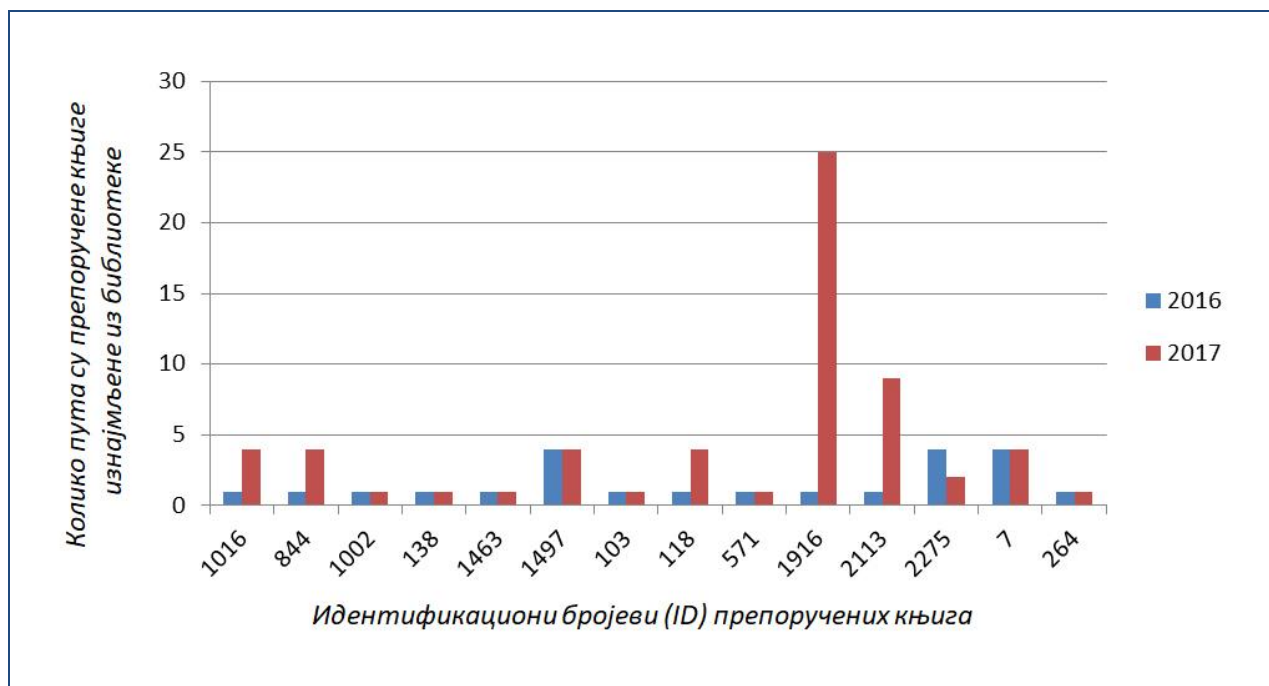
```
SELECT recommender.book_id, book.name,
COUNT(recommender.book_id) OVER (PARTITION BY recommender.book_id ORDER BY
recommender.book_id ROWS BETWEEN UNBOUNDED PRECEDING AND UNBOUNDED
FOLLOWING)
FROM recommender JOIN book ON (recommender.book_id=book.count)
WHERE recommender.book_id IN ('1016', '844', '1002', '138', '1463', '1497', '103', '118', '571',
'1916', '2113', '2275', '7', '264')
AND recommender.date_taken>'2017-1-1' AND recommender.date_return<'2017-12-1';
```

Резултати извршеног упита су приказани у прилогу Б3.

Упоредни приказ препоручених књига изнајмљених из библиотеке у 2016. и 2017. години је приказан у табели б.

Табела 6. Упоредни приказ изнајмљених књига у 2016. у односу на 2017. годину

ID књига	Наслови књига (препоручене књиге студентима применом <i>Hadoop</i> екосистема)	Изнајмљене из библиотеке (2016) (број пута)	Изнајмљене из библиотеке (2017) (број пута)
1016	Основи електронике и телекомуникација	1	4
844	Хидраулика: увод са примерима управљања	1	4
1002	Моторна возила I: општи и теоријски део	1	1
138	Електричне машине: за трећи разред електротехничке школе	1	1
1463	Комуторни мотори	1	1
1497	Термоелектране	4	4
103	Физика: механика чврстих	1	1
118	Трансформатори	1	4
571	Психолошки услови трансфера учења	1	1
1916	Енергетски трансформатори и генератори	1	25
2113	Фотонапонска постројења: планирање	1	9
2275	Соларне технологије: топлотни и фотоелектрични системи	4	2
7	Основи рачунарске технике	4	4
264	Математички приручник	1	1



Слика 49. Колико пута су препоручене књиге изнајмљене из библиотеке

Као што је приказано на слици 49, у 2016. години књиге које су узете за евалуацију система су изнајмљене 23 пута, у поређењу са 2017. годином, када су исте књиге изнајмљене 62 пута, што је приближно 269,5% више.

Друга евалуација решења је спроведена путем онлајн упитника за процену имплементираног прототипа екосистема. Упитник је интегрисан у *LMS Moodle* систем за управљање учењем у Високој школи електротехнике и рачунарства струковних студија у Београду. Упитник је коришћен као метода за утврђивање да ли је препорука заснована на *Hadoop*-у са интегрисаним подацима из више извора прикладнија за студенте (на основу њихових утисака) од система препоруке електронске продавнице књига образовне институције и примењеног једног скупа података. Узорак обухвата 220 студената треће године студија, студијског програма Електронско пословање током зимског семестра 2016/17. године. Сви студенти су пристали да учествују у истраживању. Упитник садржи десет питања. Свако питање је подржано *Moodle* системом за управљање учењем: одговори са више избора и тачно / нетачно (енгл. *true-false*). Два важна питања (четврто и десето) су заснована на Ликертовој скали од пет тачака (енгл. *Five-point Likert scale*). Онлајн упитник је приказан у табели 7.

Табела 7. Онлајн упитник за прикупљање података од студената на основу њихових утисака

Редни број	Питање
1	Да ли изнајмљујете књиге из библиотеке школе? (а) Да; (б) Не
2	Током школске 2016/17. године, колико књига сте изнајмили? Зависи од претходног питања: (а) < 5 ; (б) ≥ 5
3	Да ли на <i>LMS Moodle</i> платформи добијате препоруку литературе? (а) Да; (б) Не
4	Оцените добијену препоруку на <i>LMS Moodle</i> платформи (уколико је одговор на треће питање: Да): (а) Потпуно неодговарајућа; (б) Донекле неодговарајућа; (в) Неутрална; (г) Донекле одговарајућа; (д) Потпуно одговарајућа
5	На основу изабраних предмета у школској 2016/17. години, да ли имате понуђене е-књиге за преузимање на Вашем студентском сервису информационог система школе? (а) Да; (б) Не
6	Колико Вам је е-књига доступно за преузимање на Вашем студентском сервису? Зависи од претходног питања: (а) < 5 ; (б) ≥ 5
7	Да ли сте се регистровали и обавили куповину на електронској продавници књига школе? (а) Да; (б) Не
8	Колико књига сте купили на електронској продавници књига школе? Зависи од претходног питања: (а) < 5 ; (б) ≥ 5
9	Са новим приступом и пријавом на систем електронске продавнице књига, да ли добијате препоруку? (а) Да; (б) Не
10	Оцените добијену препоруку на електронској продавници књига (уколико је одговор на седмо питање: Да): (а) Потпуно неодговарајућа; (б) Донекле неодговарајућа; (в) Неутрална; (г) Донекле одговарајућа; (д) Потпуно одговарајућа

5.4. АНАЛИЗА ПОСТИГНУТИХ РЕЗУЛТАТА

Прво разматрање постигнутих резултата је *big data* аналитика која илуструје потенцијалне користи од примене система препоруке паметне библиотеке са интегрисаним вишеструким изворима различитих података у контексту задовољавања потреба студената и економских користи организације, и укупним задовољством пружених услуга кроз управљање ИТ пројектима, чиме се промовише међусобно разумевање и дугорочно поверење (*Liu et al.*, 2017). То потврђују резултати који показују пораст од приближно 269,5% изнајмљених, релевантних књига које су највише препоручиване студентима путем система препоруке паметне библиотеке заснованог на *big data* технологијама, препоручених путем система за

управљање учењем *LMS Moodle* током тестирања прототипа модела.

Друго разматрање се односи на регистрацију и куповину путем електронске продавнице књига образовне институције. Резултати су показали да се 40% од укупног броја студената који су учествовали у онлајн упитнику нису регистровали и нису обавили куповину на електронској продавници књига образовне институције. Без регистрације и обављене куповине, систем није био у могућности да одреди корисничке преференције и стога није могао да изгради модел корисника и да генерише препоруку због проблема са хладним стартом (*Schein et al.*, 2002). Та група студената је искључена из коначних резултата анализе система препоруке паметне библиотеке засноване на *Hadoop* екосистему.

Остали студенти, учесници анкетног упитника су испуњавали услове за анализу и евалуацију упитника.

Узорак у анализи и евалуацији постигнутих резултата укључује студенте:

- који су имали приступ електронским књигама за преузимање са својих онлајн студентских сервиса путем информационог система образовне институције;
- који су обавили регистрацију и куповину на електронској продавници књига образовне институције; и
- који су изнајмили књиге из библиотеке образовне институције.

Резултати упитника (приказани у табели 8) показују да, на основу анкетираних учесника са исказом њихових утисака, и резултатима анализе анкете, са чиниоцима где је:

- \bar{x} средња оцена (енгл. *Mean grade*), приказана у вредности од 1 до 5; и
- δ стандардно одступање (енгл. *Standard deviation*),

систем паметне библиотеке, применом различитих података интегрисаних из више извора, генерише одговарајуће препоруке, иако постоје мале разлике:

- 24,6% студената је оценило добијену препоруку модела паметне библиотеке заснованог на *big data* технологијама путем *LMS Moodle* платформе у потпуности одговарајућом, са средњим оценом од 3,38 и стандардним одступањем од 1,25;
- док је 21,1% студената оценило препоруку електронске продавнице књига образовне институције у потпуности одговарајућом са средњом оценом од 3,17 и стандардним одступањем од 1.28.

Одговори на питања која се тичу система препоруке, показују да студенти, учесници анкете, подржавају препоруку садржаја применом модела паметне библиотеке заснованог на *big data* технологијама и *Hadoop* оквира интегрисаног у систем за управљање учењем *LMS Moodle*.

Табела 8. Дескриптивна статистика ($n=132$) студената базирана на основу њихових утисака, са приказом средње оцене и стандардним одступањем, везаним за питања под редним бројем 4 и редним бројем 10

Питање	Потпуно одговарајућа (%)	Донекле одговарајућа (%)	Неутрална (%)	Донекле неодговарајућа (%)	Потпуно неодговарајућа (%)	\bar{x}	δ
Оцените добијену препоруку на <i>LMS Moodle</i> платформи	24.6	22.8	26.3	19.3	7.0	3.38	1.25
Оцените добијену препоруку на електронској продавници књига	21.1	19.3	24.6	26.3	8.8	3.17	1.28

6. ДИСКУСИЈА

Из перспективе друштва у којем данас живимо и цивилизације која је заснована на информацијама, нове информационе технологије омогућавају библиотекама транзицију традиционалне фазе аутоматизације пословних процеса и изградњу дигиталне библиотеке (*Li et al.*, 2019). Очекује се да ће се иновације нових технологија убрзано развијати док библиотеке не буду спремне да одговоре на нове технолошке изазове као што су услуге мобилног Интернета и вештачка интелигенција (*Wang*, 2017). Из ове перспективе, велики подаци и *big data* аналитика ће неминовно имати дубок утицај на услуге које пружају савремене дигиталне библиотеке. У складу са тим, у изградњи функционалне и оперативне паметне библиотеке приказане у докторској дисертацији, примена нових технолошких достигнућа је неминовност, и постепено добија све већу пажњу научно-истраживачке заједнице.

Међународна федерација библиотечких удружења и институција (енгл. *International Federation of Library Associations and Institutions – IFLA*) сврстава вештачку интелигенцију као један од четири главна технолошка тренда и тврди да вештачка интелигенција има три главне импликације у односу на будућност развоја библиотечких информационих система (*IFLA*, 2017):

- 1) прегледачи нове генерације који се проширују изнад претраживања кључних речи и семантичке анализе Интернет садржаја;
- 2) интегрисано препознавање говора и аутоматизовани вишејезични машински превод у реалном времену; и
- 3) идентификација услуга рачунарства у облаку за комплексне и различите Интернет садржаје.

Стога, образовне институције треба да усвоје утицај који нове технологије, попут великих података, машинског учења и вештачке интелигенције, имају на развој библиотечких информационих система. На основу анализе литературе и научних достигнућа, у дисертацији је предложен модел паметне библиотеке заснован на *big data* технологијама са будућим смерницама за наредну фазу развоја паметног библиотечког система, и важном улогом коју персонализоване информационе услуге имају.

Појавом великих података идентификују се нови пословни захтеви који могу довести до открића нових знања. Прилагођене и персонализоване услуге корисницима, склоне честим променама током времена коришћења система, издвајају вредност која привлачи пажњу. Потенцијал и ефикасност доношења одлука заснованих на подацима континуирано јача, а профит од анализе великих података се остварује. Потражња за великим подацима расте у домену дигиталних библиотека (*De Mauro et al.*, 2016). Међутим, релативно мало студија је истраживало дигиталне библиотеке у односу на велике податке. Један од разлога за недостатак оваквих студија је тај што постоји мишљење да традиционални системи за

управљање базама података могу да одговоре свакодневним захтевима за складиштењем података и њиховом пословном обрадом у дигиталним библиотечким системима (*Xu et al.*, 2017). Примера ради, *Hessman* (2013) сматра да су дистрибуирани системи непотребни (*Li et al.*, 2019). Међутим, већина традиционалних информационих система није у стању да складишти, обрађује и анализира велике податке. Чак и ако се такви системи редовно ажурирају, неће бити у стању да процесирају растуће количине података због све већег броја корисника, доступног садржаја, веб локација, укупног броја посетилаца и обима њихове употребе, са убрзаним трендом раста у будућности (*Chen et al.*, 2015). Због тога, као важна база људског знања, библиотеке постепено чувају све веће количине података. Не само да је величина, врста и брзина стварања података досегла висок ниво, већ су интеграција и свеобухватна употреба спољних извора великих података постали кључни за побољшање нивоа услуга у постојећим библиотекама. Због тога се очекује да паметне библиотеке имају све важнију улогу у анализи великих података и коришћењу информација.

У поређењу са другим доменима, истраживање великих података у односу на библиотеке је ограничено. Укупна конкурентност у истраживању је углавном мала; истраживачки напори су местимични; недостаје емпиријско истраживање; и нема довољно нагласка на практичној примени великих података у библиотекама (*Jing*, 2014; *Li et al.*, 2019). Примера ради, одређена истраживања указују да су системи дигиталних библиотека сасвим довољни у односу на информациони систем и организационо управљање компаније, и често не узимају у обзир нове технологије (*Li et al.*, 2019). *Horstmann and Witt* (2017) тврде да разлози могу бити повезани са ограниченим буџетом, иако је тешко утврдити да ли ће промене, као што су опадање трошкова хардверских компоненти и континуирани раст употребљивости софтвера, ублажити потешкоће које су уско повезане са изградњом савременог библиотечног информационог система (*Wang et al.*, 2016). *Lee* (2013) анализира питања безбедности протока информација и заштите приватности личних података у примени нових технологија, које неминовно треба да имају велики утицај на свеобухватније интеграције великих скупова података и у библиотечким системима. Напредак у области библиотечног пословања је релативно спор у погледу прихватања и асимилације стручних знања и нових технологија. За дигиталну библиотеку, велики број информационих технологија које се брзо развијају поопштрава овај проблем. Библиотеке чак и у фази дигитализације прихватају суштински изоловане дигиталне стратегије, што је још један разлог зашто је развој библиотека склон конзервативном пословању (*Jianzhong*, 2017). Традиционално библиотечно пословање се превасходно фокусира на традиционалну литературу, а не на информативни садржај. Оперативан модел библиотека и даље остаје зависан о традиционалном систему који је заснован на комерцијалном објављивању. Већина дигиталних библиотека углавном се фокусира на дигитализацију докумената, организацију и складиштење дигиталних докумената, проналажење и достављање докумената. Неспорно је да традиционалан модел има вредност. Међутим, превелико ограничавање ресурса и услуга доводи у опасност будућност библиотеке (*Xiaolin*, 2011). Са широком употребом Интернет технологије и мобилних уређаја, библиотека више није једина организација информационих сервиса и база знања. Традиционалан и конзервативан систем управљања библиотекама постепено слаби њену основну конкурентност и води ка Ланкастеровој дискусији и „теорији о изумирању библиотеке“ (*Xianchun and Ming*, 2014; *Lancaster*, 1982).

Иако *big data* технологије, не само да омогућавају средства и могућност акције, већ

обезбеђују потпуно нов начин за прибављање чињеница за доношење одлука. Уколико се прикупљени подаци правилно користе и успешно анализирају, постоји већа вероватноћа стварања знања од веће вредности. Значај успешне примене великих података и нових технологија осигурава конкурентност организација, ефикасност одлучивања и доводи до неизбежних промена у пословним процесима. *Li et al.* (2019) емпиријски откривају значај анализе постојећих библиотечких података, као што су информације о понашању корисника (историја прегледања, претраживања, преузимања, задуживања и др.) са подацима друштвених медија, што може постићи свеобухватну анализу навика читања корисника у библиотеци, корисничке интеракције са системом и употребу мреже. Применом техника рударења података (енгл *Data mining*), издавачи могу одредити које ће књиге објавити у складу са историјом прегледа и преузимања електронских књига читалаца, и да дистрибутери могу користити ове информације за аутоматско одређивање понуда за куповину, прављење библиографија које препоручују читаоци из перспективе корисника и организовање маркетиншких активности из перспективе издавача (*Li et al.*, 2019). Са појавом великих података, све већи број библиотека користи софистициране алате, као што су анализа перформанси учења и анализа перформанси научних истраживања у поређењу са традиционалним методама анализе података, попут статистичких метода (*Cox and Jantti*, 2012). *Xinning* (2015) анализира нормативе за реформу библиотечког пословања у *big data* окружењу у односу на три аспекта: (1) изградња потребних ресурса – проширивање обима интегрисаних извора и интеракције података; (2) примена нових технолошких достигнућа – примена семантичке анализе са техникама проналажења образаца података; и (3) пружање сервиса и услуга – пребацивање услуга са заједничког пасивног модела на проактивнији, аутоматски и персонализован модел. Међутим, анализа научне литературе указује на недостатак истраживања у практичној примени *big data* технологија у изградњи паметног библиотечког информационог система способног за *big data* анализу различитих типова података из више независних извора и аутоматизацију пословних процеса који доприносе већем профиту организације и повећавају задовољство корисника препоручивањем садржаја од интереса.

Лапкинов извештај објављен у Гартнеру (2012) истиче да: „*Big data аналитика се не своди само на MapReduce и Hadoop системе. Иако многе организације сматрају да су ове дистрибуиране технологије обраде података једине релевантне big data технологије, постоје алтернативе*“. Те алтернативе система база података за паралелну обраду и анализу (нпр. *Vertica, Teradata, Netezza, SQL Server, Greenplum, ParAccel*) су скупе, тешке за администрацију, имају недостатак толеранције на грешке и брзине обраде великих упита (*Pavlo et.al.*, 2009; *Sakr*, 2016). У пракси, *Hadoop* као пројекат отвореног кода постигао је велики успех, са све већим моментом у истраживању и развоју у образовним и пословним доменама (*Simović*, 2018). Ова технологија, са сервисима, инфраструктуром и модулима од значаја; њиховим комбиновањем у сврху прикупљања, складиштења и анализе великих скупова података, може да омогући чак и малим компанијама да постигну конкурентску предност у условима хиперконкурентности и слободне тржишне економије. *Hadoop* као софтвер отвореног кода омогућава обраду велике количине података лако и исплативо (*Sakr*, 2016). Са друге стране, организације треба да се прилагоде захтевима новог времена; да ангажују кадар који је школован за овакву врсту посла, и да успоставе пословни процес који је у стању да у потпуности искористи предности *big data* аналитике. Лоше идентификовани улазни подаци даће нетачне излазне резултате без обзира на квалитет *big data* решења. Један

од главних изазова приликом изградње паметне библиотеке је пројектовање и реализација система способног за анализу и обраду података који се могу паралелно процесирати применом *Hadoop* екосистема. Правилна анализа вишеструких података из различитих извора интегрисаних у паметну библиотеку није корисна само за повећање ефикасности пословних процеса организације и економске користи, већ је ефикасна и за додатно мотивисање корисника препоруком садржаја од интереса (*Simović, 2018*).

Добијени резултати показују да предложени систем позитивно утиче на параметре перформанси библиотеке повећавајући број изнајмљених књига и перцепцију студената, што указује на то да је такав систем посебно погодан за примену у образовним институцијама као саставни део континуираног процеса учења. Користећи предложени систем за анализу података, библиотекари и библиотечки системи могу постићи *big data* операције паметне библиотеке за долазеће генерације.

6.1. ОГРАНИЧЕЊА

Big data аналитика захтева ангажовање научника података (*Davenport and Patil, 2012*) чија се улога значајно мења у библиотечким пословним системима. Да би се постигла конкурентска предност у *big data* окружењу, библиотеке треба да имплементирају одговарајући *big data* оквир са сервисима и модулима од значаја који ће допринети задовољству корисника и остварити дугорочно поверење (*Ratledge and Sproles, 2017*).

Докторска дисертација представља практичну примену *big data* система препоруке и први корак стварања потпуно оперативног система паметне библиотеке. Извршена је интеграција и обрада четири независна извора података. Добијени резултати су визуализовани и евалуирани. Имплементирана архитектура система има ограничења и недостатке када су у питању дистрибуирани системи великих размера. Захевне апликације за рад на високо скалабилним кластерима рачунара су серверски токови података (енгл. *Server-side flow*), оптимизација приступа подацима (*Ishii and de Mello, 2012*) и побољшане перформансе за репликацију података, дистрибуцију, миграцију и приступ паралелизму (*Philip Chen and Zhang, 2014*).

Да би се побољшале перформансе система и искористио пун потенцијал предложеног *big data* модела паметног библиотечког екосистема, неопходна је будућа интеграција са вишеструким подацима из других извора (подаци *IoT* уређаја и подаци са друштвених мрежа) (*Simović, 2018*). Исто тако, проблематика проређених података (енгл. *Data sparsity*) (*Guo et al., 2014*) остаје питање.

6.2. ИСТРАЖИВАЧКЕ ИМПЛИКАЦИЈЕ

Инфраструктура паметног библиотечког екосистема, на нивоу скупа података, има могућност управљања и релационим и нерелационим структурама, омогућавајући системе за подршку одлучивању и обраду података у реалном времену коришћењем сервиса и модула за

интероперабилност интерних и екстерних извора. Такође пружа аналитичке апликације за анализу предвиђања и интерактивно истраживање обједињених података из различитих извора, обезбеђујући притом елементе персонализације са циљем издвајања вредности.

Способност система да одговори персонализованим захтевима корисника у *big data* окружењу, чине три кључне импликације са следећим карактеристикама:

- 1) Експоненцијалан раст укупне количине података отежава добијање информација од значаја. Проблем преоптерећења информација постаје израженији. Стога је издвајање садржаја из великих података и различитих извора, који за корисника може бити од интересне вредности, кључно за побољшање задовољства корисника и већег поверења у паметни библиотечки информациони систем.
- 2) Све већа количина података доводи до све већих веза података. Такве везе не само да могу унапредити разумевање информација и олакшати начине проналажења циљаних образаца ефикасније и ефективније, већ такође пружају неопходне и основне услове за даља истраживања и анализу скривених вредности које традиционални ресурси не могу да пруже. У великој количини података постоји велики број асоцијација међу подацима, попут асоцијација друштвених података корисника, асоцијација група корисника, асоцијација корисника и ресурса и асоцијација различитих ресурса (*Li et al.*, 2019). Такве везе омогућавају корисницима да лакше и брже добију интересни садржај. Поред тога, такве везе могу да генеришу нове информационе моделе корисника и могу се применити за стварање кредибилних информационих услуга комбиновањем постојећих образаца модела корисника.
- 3) Корисници, својим приступом и прегледом апликације, добијају податке и интересни садржај анализом и интеракцијом са информационим системом. Разумевање и анализу садржаја одређују интегрисани подаци из различитих извора, специфичног окружења, тренутних корисничких захтева и интеракције семантичких апликативних односа различитих система. Контекст тако генерисаних персонализованих услуга имаће значајан утицај на разумевање добијених података од стране корисника.

Омогућавањем персонализованих услуга, савремени библиотечки системи значајно могу повећати разноликост корисничких услуга и пружити корисницима одговарајуће информационе ресурсе и садржаје, што смањује трошкове корисничког времена, омогућава аутоматизацију пословних процеса и повећава добит везану за организовање библиотечких информација. Да би задовољили све већу потражњу корисника за персонализованом услугом, потребно је да савремени библиотечки системи усвоје примену нових технолошких достигнућа и *big data* технологија за оптимизацију пројектовања библиотечког екосистема.

Практична апликација модела паметне библиотеке заснованог на *big data* технологијама као савременог библиотечког екосистема обухвата следеће елементе конструкције:

- Примену скупова података традиционалних библиотечких система који обухватају: податке штампане литературе; колекције дигиталних ресурса; и ресурсе релационе базе података.

- Примену великих података који унапређују услове складиштења и употребу постојећих извора података.
- Интеграцију различитих ресурса великих података у постојеће системе и обогаћивање постојеће величине и врсте података.
- Побољшање традиционалних система интеграцијом ново-генерисаних података различитих формата са постојећим ресурсима библиотечког система.
- Омогућавање нових облика и метода пружања услуга корисницима.
- Унапређење традиционалне технолошке платформе која се може побољшати технологијом која је потребна за обраду великих података, као што је *big data* технологија прикупљања, складиштења, анализе и рударења података.
- Директно или индиректно омогућавање корисницима приступ ресурсима података.
- Идентификације појединачних образаца интереса корисника тако да се услуге могу прилагодити променљивим информационим потребама.
- Развијање персонализованог режима услуге један на један (енгл. *one-to-one*), у односу на традиционалан режим услуге један на много (енгл. *one-to-many*).
- Пружање проактивних услуга, попут персонализованих препорука у складу са интересовањима корисника.
- Могућност истовременог приступа корисника интегрисаним системима са више уређаја ради побољшања услуга у свим аспектима.
- Визуализацију корисничког приступа услугама система паметног библиотечког система на интуитиван и практичан начин.

Платформа паметне библиотеке подржава истраживаче и научнике за анализу података за управљање одлукама у реалном времену. Нова технолошка решења, као што су расподељени и дистрибуирани оквири, паралелно процесирање, велики подаци и вештачка интелигенција, стални је темељ иновација у савременој паметној библиотеци.

6.3. ПРАКТИЧНЕ ИМПЛИКАЦИЈЕ

Брзо ширење и експоненцијалан раст доступних података из различитих извора створили су нове приступе њиховој обради и нови канал за свеобухватан преглед информација од значаја у библиотечким системима. *Hadoop* екосистем је оквир који решава проблеме управљања великим подацима у паметним окружењима. *Big data* технологије имају фундаменталан значај у управљању информацијама у контексту откривања знања и побољшања библиотечких пословних процеса.

Предложени екосистем има способност да побољша библиотечке операције из неколико

разлога:

- има способност прикупљања података из више различитих извора;
- одликује се поузданошћу са толеранцијом на грешке у односу на традиционалне системе управљања подацима;
- има способност чувања личних преференција и корисничких атрибута важних за даље прогнозе и предикције;
- анализира садржај различитих података са циљем проналажења сличности из више извора;
- решава проблем у контексту истовременог приступа вишеструким подацима;
- скалира ресурсе који управљају већом количином података;
- анализира и обрађује податке применом *big data* технологија;
- има способност дељења увида у знање широм целокупног информационог система организације.

Практичне импликације доприносе и задовољству корисника (*Shen et al., 2013*) препорученим садржајем добијеним путем система препоруке паметне библиотеке у *big data* окружењу, који не само да студентима и корисницима пружа корисне и високо персонализоване садржаје, већ изграђује дугорочно поверење стварањем листе препорука са већом прецизношћу.

7. НАУЧНИ И СТРУЧНИ ДОПРИНОСИ

Најзначајнији допринос докторске дисертације је развој модела паметне библиотеке заснованог на *big data* технологијама који се може применити у образовним институцијама са циљем остваривања бољих пословних резултата и повећања задовољства студената у континуираном процесу учења. Модел је заснован на предлагању садржаја од интереса за студенте путем система препоруке интеграцијом више независних извора разноврсних података у паметни библиотечки информациони систем.

Кључни научни доприноси докторске дисертације се огледају у:

- формалном опису модела паметне библиотеке заснованог на *big data* технологијама;
- анализи метода и развоју модела система препоруке у *big data* окружењу паметне библиотеке;
- развоју модела *big data* инфраструктуре и архитектуре паметне библиотеке;
- моделу интеграције више независних извора података у библиотечки информациони систем образовне институције;
- унапређењу образовног процеса препоручивањем релевантног библиотечног садржаја и развојем нових студентских сервиса заснованих на *big data* технологијама;
- развоју метода за оцену перформанси предложеног модела.

Докторска дисертација резултује и низом стручних доприноса од којих су најважнији:

- анализа примене *big data* технологија и система препоруке у паметним библиотекама;
- преглед и анализа технологија потребних за имплементацију система препоруке паметне библиотеке у *big data* окружењу;
- могућности интеграције инфраструктуре система препоруке паметне библиотеке са постојећим информационим системом образовне институције;
- идентификација специфичних детаља и захтева разноврсних скупова података за развој интегрисаног система препоруке у циљу побољшања резултата рада библиотека;
- анализа и валидација модела кроз економске, психолошке и технолошке перформансе система;
- могућност коришћења резултата истраживања од стране других образовних институција за имплементацију модела паметне библиотеке.

Са становишта друштвене корисности, резултати истраживања могу имати вишеструке импликације:

- резултати истраживања ће помоћи да се анализира проблем примене *big data* технологија у циљу унапређења информационог система образовне институције;
- афирмацију увођења интегрисаног система препоруке паметне библиотеке у образовне институције, заснованог на *big data* технологијама, како би се повећала мотивисаност студената у континуираном образовном циклусу;
- утврђивање потенцијала за унапређење пословних процеса образовне институције применом *big data* технологија;
- резултати истраживања ће помоћи да се анализира допринос примене *big data* технологија и система препоруке паметне библиотеке као фактора пословне ефикасности образовне институције и унапређења пословних резултата.

Резултати рада на докторској дисертацији објављени су у истакнутом међународном часопису *Library Hi Tech* у области информационих и библиотечких наука – *Information Science & Library Science*, који је на *SCI* листи, са импакт фактором 1.256 за 2018. годину, кога издаје *Emerald Publishing*:

Simović A. (2018), *A Big Data smart library recommender system for an educational institution*, *Library Hi Tech*, Vol. 36, Issue: 3, pp. 498-523, doi: 10.1108/LHT-06-2017-0131.

Резултати истраживања реализованих у оквиру докторске дисертације објављени су у више научних радова и саопштени на више међународних и домаћих скупова и конференција и то:

Simović, A. (2017), *Recommender systems in the big data environment using Mahout framework*, In *Telecommunication Forum TELFOR 2017*, Vol. 25, pp. 1-4, IEEE, doi: 10.1109/TELFOR.2017.8249472, Belgrade, Serbia.

Simović A. (2017), *Big Data Analytics in Public Administration*, XLIV Symposium on Operational Research SYM-OP-IS 2017, Vol. 44, pp. 382-387, Zlatibor, Srbija.

Simović A., Ćirović Z. (2016), *Sistem preporuke u elektronskoj trgovini u Big Data okruženju*, *ETLAN 2016*, Vol. 60, pp. RT4.5.1-4, Zlatibor, Srbija.

Simović A., Ćirović Z. (2016), *Primena Hadoop okvira u analizi velikih skupova podataka*, XV међународни научно-стручни Симпозијум INFOTEH, Vol. 15, pp. 593-598, Jahorina.

Simović, A. (2014), *Sistemi preporuke u e-trgovini*, *Sinteza 2014-Impact of the Internet on Business Activities in Serbia and Worldwide*, pp. 846-852, Belgrade, Serbia.

8. БУДУЋА ИСТРАЖИВАЊА

Имплементација модела паметне библиотеке са интегрисаним *IoT* технологијама и подацима прикупљених са друштвених мрежа захтева даља истраживања у надоградњи реализованог *big data* екосистема, са фокусом на:

- интеграцију података са платформи друштвених мрежа и аналитике у реалном времену;
- интеграцију модела паметне библиотеке са информационим системом образовне институције у циљу развоја и изградње потпуно оперативног система паметне библиотеке;
- интеграцију података са *IoT* уређаја и дистрибуираних сензора у библиотеци образовне институције и њихово даље процесирање у реализованом *big data* моделу паметне библиотеке на најефикаснији начин.

Технолошки изазови услед последица дигитализације, ширења Интернета и изузетно брзог напретка нових технологија, мењају начин образовног и пословног процеса, и намећу потребу за будућа системска решења шире примене реализованог модела паметне библиотеке у образовним институцијама. Да би се у потпуности обезбедила примарна друштвена улога библиотеке, као централног места за учење, колаборацију, размену знања и друштвени напредак, неопходно је у будућим истраживањима надоградње реализованог модела обезбедити стандарде за прикупљање и интеграцију података екстерних библиотека и других образовних установа.

Крајњи циљ будућег развоја паметног библиотечког екосистема је потпуна аутоматизација система кроз истраживање нивоа доступности и могућност интеграције нових технолошких достигнућа, попут виртуалне стварности и вештачке интелигенције.

9. ЗАКЉУЧАК

У раду је анализирана примена *big data* технологија за развој паметне библиотеке, као и могућност интеграције *big data* сервиса паметне библиотеке са информационим системом образовне институције. Дефинисане су кључне карактеристике и архитектура развијеног модела. Циљ истраживања се постиже кроз интеграцију више независних извора података у паметни библиотечки информациони систем образовне институције како би се омогућила квалитетна препорука садржаја студентима. Детаљно је приказан *big data* технолошки оквир и развијени систем препоруке садржаја у паметним библиотекама.

У оквиру дисертације пројектован је и реализован модел паметне библиотеке заснован на *big data* технологијама. У оквиру евалуације модела реализовано је тестирање и мерење релевантних параметара који утичу на ефикасност развијеног система. Економске, технолошке и образовне перформансе модела су приказане, и анализиран је утицај примене модела на задовољство корисника.

Циљ истраживања је развој и имплементација модела паметне библиотеке за управљање различитим подацима прикупљених из више извора применом *Hadoop* екосистема, који је посебно погодан за примену у образовним институцијама. Изградња система препоруке паметне библиотеке у *big data* окружењу ствара нову вредност кроз јединствене функције управљања библиотечким процесима, већим задовољством корисника и поверења у систем.

10. ЛИТЕРАТУРА

- [1] Adomavicius, G. and Tuzhilin, A. (2011), "Context-aware recommender systems", In *Recommender systems handbook*, pp. 217-253, Springer, Boston, MA.
- [2] Adomavicius, G. and Tuzhilin, A. (2005), "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17, No. 6, pp. 734-749.
- [3] Ahrens, J., Hendrickson, B., Long, G., Miller, S., Ross, R. and Williams, D. (2011), "Data-intensive science in the US DOE: case studies and future challenges", *Computing in Science and Engineering*, Vol. 13, No. 6, pp. 14-24.
- [4] Aittola, M., Ryhänen, T. and Ojala, T. (2003), "SmartLibrary–location-aware mobile library service", In *International Conference on Mobile Human-Computer Interaction*, pp. 411-416, Springer, Berlin, Heidelberg.
- [5] Amato, F., Moscato, V., Picariello, A. and Piccialli, F. (2017), "SOS: A multimedia recommender System for Online Social networks", *Future Generation Computer Systems*, Vol. 93, pp. 914-923.
- [6] Andersen, J. (2002), "Communication technologies and the concept of knowledge organization-a medium-theory perspective", *Knowledge organization*, Vol. 1, pp. 29-39.
- [7] Andersen, T. G., Bollerslev, T., Diebold, F. X. and Vega, C. (2003), "Micro effects of macro announcements: Real-time price discovery in foreign exchange", *American Economic Review*, Vol. 93, No. 1, pp. 38-62.
- [8] Aravinth, M. S., Shanmugapriya, M. S., Sowmya, M. S. and Arun, M. E. (2015), "An efficient HADOOP frameworks SQOOP and ambari for big data processing", *International Journal for Innovative Research in Science and Technology*, Vol. 1, No. 10, pp. 252-255.
- [9] Arlitsch, K. and Newell, B. (2017), "Thriving in the age of accelerations: a brief look at the societal effects of artificial Intelligence and the opportunities for libraries", *Journal of Library Administration*, Vol. 57, No. 7, pp. 789-798.
- [10] Bae, K. J., Jeong, Y. S., Shim, W. S. and Kwak, S. J. (2007), "The ubiquitous library for the blind and physically handicapped – A case study of the LG Sangnam Library, Korea", *IFLA journal*, Vol. 33, No. 3, pp. 210-219.
- [11] Bagchi, S. (2015), "Performance and quality assessment of similarity measures in collaborative filtering using mahout", *Procedia Computer Science*, Vol. 50, pp. 229-234.
- [12] Bailey, C. W. (1991), "Intelligent library systems: artificial intelligence technology and library automation systems", *Advances in Library Automation and Networking*, Vol. 4, pp. 36-37.

- [13] Balabanović, M. and Shoham, Y. (1997), "Fab: content-based, collaborative recommendation", *Communications of the ACM*, Vol. 40, No. 3, pp. 66-72.
- [14] Barbier, G. and Liu, H. (2011), "Data mining in social media", In *Social network data analytics*, pp. 327-352, Springer, Boston, MA.
- [15] Barga, R., Fontama, V. and Tok, W. H. (2015), "Cortana analytics", In *Predictive Analytics with Microsoft Azure Machine Learning*, pp. 279-283, Apress, Berkeley, CA.
- [16] Barnaghi, P., Wang, W., Henson, C. and Taylor, K. (2012), "Semantics for the Internet of Things: early progress and back to the future", *International Journal on Semantic Web and Information Systems IJSWIS*, Vol. 8, No. 1, pp. 1-21.
- [17] Bekkerman, R., Bilenko, M. and Langford, J. (2011), "Scaling up machine learning: Parallel and distributed approaches", *Cambridge University Press*.
- [18] Beliakov, G., Calvo, T. and James, S. (2011), "Aggregation of preferences in recommender systems", In *Recommender systems handbook*, pp. 705-734, Springer, Boston, MA.
- [19] Bell, S. J. and Shank, J. (2004), "The blended librarian: A blueprint for redefining the teaching and learning role of academic librarians", *College & Research Libraries News*, Vol. 65, No. 7, pp. 372-375.
- [20] Bernardes, D., Diaby, M., Fournier, R., FogelmanSoulié, F. and Viennet, E. (2014), "A social formalism and survey for recommender systems", *ACM SIGKDD Explorations Newsletter*, Vol. 16, No. 2, pp. 20-37.
- [21] Bertot, J. C. and Choi, H. (2013), "Big data and e-government: issues, policies, and recommendations", In *Proceedings of the 14th Annual International Conference on Digital Government Research*, pp. 1-10, ACM.
- [22] Beyer, M.A. and Laney, D. (2012), *The Importance of Big Data: A Definition*, доступно на: www.gartner.com (датум приступа: 15.5.2017).
- [23] Bhat, W. A. (2018), "Long-term preservation of big data: prospects of current storage technologies in digital libraries", *Library Hi Tech*, Vol. 36, No. 3, pp. 539-555.
- [24] Bi, Z., Liu, Y., Krider, J., Buckland, J., Whiteman, A., Beachy, D. and Smith, J. (2018), "Real-Time Force Monitoring of Smart Grippers for Internet of Things (IoT) Applications", *Journal of Industrial Information Integration*, Vol. 11, pp. 19-28.
- [25] Bizer, C., Heath, T. and Berners-Lee, T. (2009), "Linked data-the story so far", *Semantic services, interoperability and web applications: emerging concepts*, pp. 205-227.
- [26] Bogorin, A., Parditka, G. and Rempis, P. (2013), "Das Projekt Blended Library an der Universitätsbibliothek Tübingen—ein Statusbericht", *Bibliotheksdienst*, Vol. 47, No. 12, pp. 964-975.

- [27] Boulon, J., Konwinski, A., Qi, R., Rabkin, A., Yang, E. and Yang, M. (2008), "Chukwa, a large-scale monitoring system", In *Proceedings of CCA*, Vol. 8, pp. 1-5.
- [28] Boyd, D. and Crawford, K. (2012), "Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon", *Information, Communication, & Society*, pp. 662-679.
- [29] Brian, A. L. A., Arockiam, L. and Malarchelvi, P. D. S. K. (2014), "An IOT based secured smart library system with NFC based book tracking", *International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)*, Vol. 11, No. 5.
- [30] Burke, R. (2002), "Hybrid recommender systems: Survey and experiments", *User modeling and user-adapted interaction*, Vol. 12, No. 4, pp. 331-370.
- [31] Burke, R., O'Mahony, M. P. and Hurley, N. J. (2015), "Robust collaborative recommendation", In *Recommender systems handbook*, pp. 961-995, Springer, Boston, MA.
- [32] Burnap, P., Rana, O., Williams, M., Housley, W., Edwards, A., Morgan, J., ... and Conejero, J. (2015), "COSMOS: Towards an integrated and scalable service for analysing social media on demand", *International Journal of Parallel, Emergent and Distributed Systems*, Vol. 30, No. 2, pp. 80-100.
- [33] Buyya, R. and Ranjan, R. (2010), "Special section: Federated resource management in grid and cloud computing systems", *Future Generation Computer Systems*, Vol. 26, No. 8, pp. 1189-1191.
- [34] Buyya, R., Calheiros, R.N. and Dastjerdi, A.V. (2016), *Big Data Principles and Paradigms*, Morgan Kaufmann.
- [35] Cantador, I., Bellogín, A. and Castells, P. (2008), "A multilayer ontology-based hybrid recommendation model", *Ai Communications*, Vol. 21, No. 2-3, pp. 203-210.
- [36] Cao, G., Liang, M. and Li, X. (2018), "How to make the library smart? The conceptualization of the smart library", *The Electronic Library*, Vol. 36, No. 5, pp. 811-825.
- [37] Capuano, N., Gaeta, M., Ritrovato, P. and Salerno, S. (2014), "Elicitation of latent learning needs through learning goals recommendation", *Computers in Human Behavior*, Vol. 30, pp. 663-673.
- [38] Chambers, S. (Ed.). (2013), *Catalogue 2.0: the future of the library catalogue*, Facet Publishing.
- [39] Chandio, A. A., Tziritas, N. and Xu, C. Z. (2015), "Big-data processing techniques and their challenges in transport domain", *ZTE Communications*, Vol 1, pp.10.
- [40] Chang, F., Dean, J., Ghemawat, S., Hsieh, W. C., Wallach, D. A., Burrows, M., ... and Gruber, R. E. (2008), "Bigtable: A distributed storage system for structured data", *ACM Transactions on Computer Systems (TOCS)*, Vol. 26, No. 2, pp. 1-26.

- [41] Chauhan, A., Fontama, V., Hart, M., Tok, W. H. and Woody, B. (2014), *Introducing Microsoft Azure HDInsight*. Microsoft press.
- [42] Chen, C. C. and Chen, A. P. (2007), "Using data mining technology to provide a recommendation service in the digital library", *The Electronic Library*, Vol. 25, No. 6, pp. 711-724.
- [43] Chen, G. D. and Chao, P. Y. (2008), "Augmenting Traditional Books with Context-Aware Learning Supports from Online Learning Communities", *Journal of Educational Technology & Society*, Vol. 11, No. 2, pp. 27-40.
- [44] Chen, H. (2017), "Applications of cyber-physical system: a literature review", *Journal of Industrial Integration and Management*, Vol. 2, No. 3, 1750012.
- [45] Chen, H. L., Doty, P., Mollman, C., Niu, X., Yu, J. C. and Zhang, T. (2015), "Library assessment and data analytics in the big data era: Practice and policies", *Proceedings of the Association for Information Science and Technology*, Vol. 52, No. 1, pp. 1-4.
- [46] Chen, L. and Zhang, H. (2016), "Smart library from the angle of its service", *Chinese Journal of Medical Library and Information Science*, Vol. 25, No. 8, pp. 58-61.
- [47] Cheng, H., Huang, L., Xu, H., Hu, Y. and Wang, X. A. (2016), "Design and implementation of library books search and management system using RFID technology", In *Intelligent Networking and Collaborative Systems (INCoS), 2016 International Conference on*, pp. 392-397, IEEE.
- [48] Chung, W. (2014), "BizPro: Extracting and categorizing business intelligence factors from textual news articles", *International Journal of Information Management*, Vol. 34, No. 2, pp. 272-284.
- [49] Civerchia, F., Bocchino, S., Salvadori, C., Rossi, E., Maggiani, L. and Petracca, M. (2017), "Industrial internet of things monitoring solution for advanced predictive maintenance applications", *Journal of Industrial Information Integration*, Vol. 7, pp. 4-12.
- [50] Cox, A. M., Pinfield, S. and Rutter, S. (2018), "The intelligent library: Thought leaders' views on the likely impact of artificial intelligence on academic libraries", *Library Hi Tech*, Vol. 37 No. 3, pp. 418-435.
- [51] Cox, B. L. and Jantti, M. H. (2012), "Discovering the impact of library use and student performance", *Academic – Papers*.
- [52] Cukier, K. and Mayer-Schoenberger, V. (2013), "The rise of Big Data: How it's changing the way we think", *Foreign Affairs*, Vol. 92, No. 3, pp. 28-40.
- [53] Cullen, K. (2005), "Delving into Data", *Library Journal*, Vol. 130, No. 13, pp. 30-33.
- [54] Da Xu, L., He, W. and Li, S. (2014), "Internet of things in industries: A survey", *IEEE Transactions on industrial informatics*, Vol. 10, No. 4, pp. 2233-2243.

- [55] Davenport, T. H. and Patil, D. J. (2012), "Data scientist", *Harvard business review*, Vol. 90, No. 5, pp. 70-76.
- [56] Davidović, B., Barać, D. and Radenković, B. (2018), "Designing a collaborative filtering recommendation system in e-commerce", *XVI Međunarodni simpozijum SymOrg 2018*, Zlatibor, Jun 2018.
- [57] Davidovic, B., Bogdanovic, Z., Djordjevic, K. and Labus, A. (2018), "Designing adaptable smart home environment based on resident's activity", *Smart Cities and Regional Development (SCRD) Journal*, Vol. 2, No. 2, pp. 51-58.
- [58] De Mauro, A., Greco, M. and Grimaldi, M. (2016), "A formal definition of Big Data based on its essential features", *Library Review*, Vol. 65, No. 3, pp. 122-135.
- [59] De Meo, P., Messina, F., Rosaci, D. and Sarné, G. M. (2017), "Combining trust and skills evaluation to form e-Learning classes in online social networks", *Information Sciences*, Vol. 405, pp. 107-122.
- [60] Dean, J. and Ghemawat, S. (2008), "MapReduce: simplified data processing on large clusters", *Communications of the ACM*, Vol. 51, No. 1, pp. 107-113.
- [61] Dede, E., Sendir, B., Kuzlu, P., Hartog, J. and Govindaraju, M. (2013), "An evaluation of cassandra for hadoop", In *Cloud Computing (CLOUD), 2013 IEEE Sixth International Conference on*, pp. 494-501, IEEE.
- [62] *Deeper insights across data with SQL Server 2016, White paper*, Microsoft Corporation, 2016, доступно на: www.microsoft.com (датум приступа: 10.1.2017).
- [63] Dent, V. F. (2007), "Intelligent agent concepts in the modern library", *Library Hi Tech*, Vol. 25, No. 1, pp. 108-125.
- [64] Despotović-Zrakić, M., Marković, A., Bogdanović, Z., Barać, D. and Krčo, S. (2012), "Providing adaptivity in Moodle LMS courses", *Journal of Educational Technology & Society*, Vol. 15, No. 1, pp. 326-338.
- [65] DeWitt, D. J., Halverson, A., Nehme, R., Shankar, S., Aguilar-Saborit, J., Avanes, A., ... and Gramling, J. (2013), "Split query processing in polybase", In *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*, pp. 1255-1266, ACM.
- [66] Diekema, A. R. (2012), "Multilinguality in the digital library: a review", *The Electronic Library*, Vol. 30, No. 2, pp. 165-181.
- [67] Du, L. and Liu, T. (2014), "Study on the development of smart library under Internet of Things", In *Applied mechanics and materials*, Vol. 529, pp. 716-720, Trans Tech Publications Ltd.
- [68] Duan, L. and Xiong, Y. (2015), "Big Data analytics and business analytics", *Journal of Management Analytics*, Vol. 2, No. 1, pp. 1-21.

- [69] Edjlali, R., Feinberg, D., Beyer, M.A. and Adrian, M. (2012), *The State of Data Warehousing in 2012*, доступно на: www.gartner.com (датум приступа: 10.5.2017).
- [70] Eliot, S. and Rose, J. (2009), *A Companion to the History of the Book*, John Wiley & Sons.
- [71] En, W. (2012), "Smart Library and the Construction of Its Service Model [J] ", *Information and Documentation Services*, Vol. 5.
- [72] Eppler, M.J. and Mengis, J. (2004), "The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines", *The information society*, Vol. 20, No. 5, pp. 325-344.
- [73] Erl, T., Khattak, W. and Buhler, P. (2016), *Big Data fundamentals: concepts, drivers & techniques*. Prentice Hall Press.
- [74] Fan, C. I. and Huang, S. Y. (2013), "Controllable privacy preserving search based on symmetric predicate encryption in cloud storage", *Future Generation Computer Systems*, Vol. 29, No. 7, pp. 1716-1724.
- [75] Fang, S., Da Xu, L., Zhu, Y., Ahati, J., Pei, H., Yan, J. and Liu, Z. (2014), "An Integrated System for Regional Environmental Monitoring and Management Based on Internet of Things", *IEEE Trans. Industrial Informatics*, Vol. 10, No. 2, pp. 1596-1605.
- [76] Feng, J., Barbosa, L. D. A. and Torres, V. (2016), *U.S. Patent No. 9,262,517, Systems and Methods for Social Media Data Mining*, Washington, DC: U.S. Patent and Trademark Office.
- [77] Fernandez, P. (2016), "Through the looking glass: envisioning new library technologies" how artificial intelligence will impact libraries", *Library Hi Tech News*, Vol. 33, No. 5, pp. 5-8.
- [78] Fey, P., Gojobori, T., Hannick, L., Hide, W., Hill, D.P., Kania, R., Schaeffer, M., St Pierre S., Twigger, S., White, O., Yrhee S.Y., Howe, D. and Costanzo, M. (2008), "Big Data: the future of biocuration", *Nature*, Vol. 455, pp. 47–50.
- [79] Gandomi, A. and Haider, M. (2015), "Beyond the hype: Big data concepts, methods, and analytics", *International journal of information management*, Vol. 35, No. 2, pp. 137-144.
- [80] Gao, B., Li, X., Liu, S. and Fang, D. (2018), "How power distance affects online hotel ratings: The positive moderating roles of hotel chain and reviewers' travel experience", *Tourism Management*, Vol. 65, pp. 176-186.
- [81] George, L. (2011), *HBase: the definitive guide: random access to your planet-size data*, O'Reilly Media, Inc.
- [82] Ghemawat, S., Gobiuff, H. and Leung, S. T. (2003), "The Google file system", *In Proceedings of the nineteenth ACM symposium on Operating systems principles*, pp. 29-43.

- [83] Golfarelli, M. and Rizzi, S. (2009), *Data warehouse design: Modern principles and methodologies*, McGraw-Hill Inc.
- [84] Gris , M.L. and Gallupe, R.B. (1999), "Information overload: Addressing the productivity paradox in face-to-face electronic meetings", *Journal of Management Information Systems*, Vol. 16, No. 3, pp. 157-185.
- [85] Grover, V., Chiang, R. H., Liang, T. P. and Zhang, D. (2018), "Creating strategic business value from big data analytics: A research framework", *Journal of Management Information Systems*, Vol. 35, No. 2, pp. 388-423.
- [86] Gundecha, P. and Liu, H. (2012), "Mining social media: a brief introduction", In *New Directions in Informatics, Optimization, Logistics, and Production*, pp. 1-17, Informs.
- [87] Guo, G., Zhang, J. and Thalmann, D. (2014), "Merging trust in collaborative filtering to alleviate data sparsity and cold start", *Knowledge-Based Systems*, Vol. 57, pp. 57-68.
- [88] Hansen, M.T. and Haas, M.R. (2001), "Competing for attention in knowledge markets: Electronic document dissemination in a management consulting company", *Administrative Science Quarterly*, Vol. 46, No. 1, pp. 1-28.
- [89] Hargittai, E. (2007), "Whose space? Differences among users and non-users of social network sites", *Journal of Computer-Mediated Communication*, Vol. 13, No. 1, pp. 276-297.
- [90] Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A. and Khan, S. U. (2015), "The rise of "big data" on cloud computing: Review and open research issues", *Information Systems*, Vol. 47, pp. 98-115.
- [91] He, C., Tang, Y., Yang, Z., Zheng, K. and Chen, G. (2014), "SRSH: a social recommender system based on Hadoop", *International Journal of Multimedia and Ubiquitous Engineering*, Vol. 9, No. 6, pp. 141-152.
- [92] Herlocker, J. L., Konstan, J. A., Borchers, A. and Riedl, J. (1999), "An algorithmic framework for performing collaborative filtering", In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 230-237, ACM.
- [93] Herlocker, J. L., Konstan, J. A., Terveen, L. G. and Riedl, J. T. (2004), "Evaluating collaborative filtering recommender systems", *ACM Transactions on Information Systems (TOIS)*, Vol. 22, No. 1, pp. 5-53.
- [94] Herlocker, J.L., Konstan, J.A. and Riedl, J. (2000), "Explaining collaborative filtering recommendations", *CSCW '00 Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pp. 241-250.
- [95] Hessman, T. (2013), Putting big data to work, *Industry Week*, Vol. 262, No. 4, pp. 14-18.

- [96] Hey, T., Tansley, S. and Tolle, K. M. (2009), *The fourth paradigm: data-intensive scientific discovery*, Redmond, WA: Microsoft research.
- [97] Hodder, I. (2013), *The meanings of things: material culture and symbolic expression*, Routledge.
- [98] Hoffman, S. (2013), *Apache Flume: distributed log collection for Hadoop*, Packt Publishing Ltd.
- [99] Holmes, A. (2012), *Hadoop in practice*, Manning Publications Co.
- [100] Horstmann, W. and Brase, J. (2016), "Libraries and Data – Paradigm Shifts and Challenges", *Bibliothek Forschung und Praxis*, Vol. 40, No. 2, pp. 273-277.
- [101] Horstmann, W. and Witt, M. (2017), "Libraries tackle the challenge of research data management", *IFLA journal*, Vol. 43, No. 1, pp. 3-4.
- [102] Hu, J. and Zhang, Y. (2016), "Chinese students' behavior intention to use mobile library apps and effects of education level and discipline", *Library Hi Tech*, Vol. 34, No. 4, pp. 639-656.
- [103] Huang, C. H. and Wang, C. M. (2011), "Usability analysis in gesture operation of interactive e-books on mobile devices", In *International Conference of Design, User Experience, and Usability*, pp. 573-582, Springer, Berlin, Heidelberg.
- [104] Hunt, P., Konar, M., Junqueira, F. P. and Reed, B. (2010), "ZooKeeper: Wait-free Coordination for Internet-scale Systems", In *USENIX annual technical conference*, Vol. 8, No. 9.
- [105] Hunt, R.E. and Newman, R.G. (1997), "Medical knowledge overload: a disturbing trend for physicians", *Health care management review*, Vol. 22, No. 1, pp. 70-75.
- [106] IFLA (2013), *IFLA statement on libraries and development*, доступно на: www.ifla.org (датум приступа: 11.11.2018).
- [107] IFLA (2017), *Advances in artificial intelligence*, доступно на: www.trends.ifla.org (датум приступа: 11.6.2020).
- [108] Ishii, R. P. and de Mello, R. F. (2012), "An online data access prediction and optimization approach for distributed systems", *IEEE transactions on parallel and distributed systems*, Vol. 23, No. 6, pp. 1017-1029.
- [109] Islam, M., Huang, A. K., Battisha, M., Chiang, M., Srinivasan, S., Peters, C., ... and Abdelnur, A. (2012), "Oozie: towards a scalable workflow management system for Hadoop", In *Proceedings of the 1st ACM SIGMOD Workshop on Scalable Workflow Execution Engines and Technologies*. ACM, pp. 1-10.

- [110] Ismail, A. S. and Al-Feel, H. (2015), "Digital library recommender system on hadoop", In *Network Cloud Computing and Applications (NCCA), 2015 IEEE Fourth Symposium*, pp. 111-114, IEEE.
- [111] Jerkov, A., Sofronijevic, A. and Stanisic, D. K. (2015), "Smart and sustainable library: Information literacy hub of a new city", In *European Conference on Information Literacy*, Springer, Cham, pp. 22-30, Springer, Cham.
- [112] Jianzhong, W. U. (2017), "Re-discussion of ten hot topics in the development of librarianship", *Journal of Library Science in China*, Vol. 4, pp. 4-17.
- [113] Jing, L. (2014), "The evaluation and prospects of big data research in library in China", *Libraly Journal*, Vol. 33, No. 1, pp. 20-25.
- [114] Johnson, L., Becker, S. A., Estrada, V. and Freeman, A. (2015), *NMC horizon report: 2015 library edition*, The New Media Consortium, pp. 1-54.
- [115] Kagie, M., Van Wezel, M. and Groenen, P. J. (2011), "Map based visualization of product catalogs", In *Recommender Systems Handbook*, pp. 547-576, Springer, Boston, MA.
- [116] Kajdanowicz, T., Indyk, W., Kazienko, P. and Kukul, J. (2012), "Comparison of the efficiency of mapreduce and bulk synchronous parallel approaches to large network processing", In *2012 IEEE 12th International Conference on Data Mining Workshops*, pp. 218-225, IEEE.
- [117] Kalee, T. (2012), *How it works: underground robot library*, доступно на: www.popsoci.com (датум приступа: 07.11.2018).
- [118] Kaske, N. K. (2004), "The ubiquitous library is here", *Portal: Libraries and the Academy*, Vol. 4, No. 2, pp. 291-297.
- [119] Khan, S., Liu, X., Shakil, K. A. and Alam, M. (2017), "A survey on scholarly data: From Big Data perspective", *Information Processing & Management*, Vol. 53, No. 4, pp. 923-944.
- [120] Khan, Z., Anjum, A., Soomro, K. and Tahir, M. A. (2015), "Towards cloud based Big Data analytics for smart future cities", *Journal of Cloud Computing*, Vol. 4, No. 1.
- [121] Kim, J. H. (2017), "A review of cyber-physical system research relevant to the emerging IT trends: industry 4.0, IoT, big data, and cloud computing", *Journal of Industrial Integration and Management*, Vol. 2, No. 03, 1750011.
- [122] Klein, C. and Kaefer, G. (2008), "From smart homes to smart cities: Opportunities and challenges from an industrial perspective", In *International Conference on Next Generation Wired/Wireless Networking*, pp. 260-260, Springer, Berlin, Heidelberg.
- [123] Klein, S. (2017), "Azure Data Lake Analytics", In *IoT Solutions in Microsoft's Azure IoT Suite*, pp. 155-172, Apress, Berkeley, CA.

- [124] Koren, Y. and Bell, R. (2015), "Advances in collaborative filtering", In *Recommender systems handbook*, pp. 77-118, Springer, Boston, MA.
- [125] Kovacevic, A., Devedzic, V. and Pocajt, V. (2010), "Using data mining to improve digital library services", *The electronic library*, Vol. 28, No. 6, pp. 829-843.
- [126] Kuilin, S. and Bo, S. (2015), "The Research and Practice of Smart Library: Taking Nanjing University Library as an Example", *New Century Library*, Vol. 7, pp. 8.
- [127] Kwok, S. K., Cheung, C. F., Lee, W. B., Tsang, A. H. and Tang, M. C. (2008), "Development of an RFID-enabled mobile smart library system", *International Journal of Enterprise Network Management*, Vol. 2, No. 2, pp. 185-197.
- [128] Lagzian, F., Abrizah, A. and Chin Wee, M. (2013), "An identification of a model for digital library critical success factors", *The Electronic Library*, Vol. 31, No. 1, pp. 5-23.
- [129] Lai, C. T., Jackson, P. R. and Jiang, W. (2017), "Shifting paradigm to service-dominant logic via Internet-of-Things with applications in the elevators industry", *Journal of Management Analytics*, Vol. 4, No. 1, pp. 35-54.
- [130] Lancaster, F. W. (1982), *Libraries and Librarians in an Age of Electronics*, Information Resources Press.
- [131] Landset, S., Khoshgoftaar, T. M., Richter, A. N. and Hasanin, T. (2015), "A survey of open source tools for machine learning with big data in the Hadoop ecosystem", *Journal of Big Data*, Vol. 2, No. 1, pp. 24.
- [132] Laney, D. (2001), "3D Data Management: Controlling Data Volume, Velocity and Variety", META group research note, Vol. 6, No. 70.
- [133] Lapkin, A. (2012), *Hype Cycle for Big Data*, доступно на: www.gartner.com (датум приступа: 12.12.2017).
- [134] Lee, J. M. (2013), "Understanding big data and utilizing its analysis into library and information services", *Journal of the Korean BIBLIA Society for library and Information Science*, Vol. 24, No. 4, pp. 53-73.
- [135] Lee, Y. J., Hosanagar, K. and Tan, Y. (2015), "Do I follow my friends or the crowd? Information cascades in online movie ratings", *Management Science*, Vol. 61, No. 9, pp. 2241-2258.
- [136] Leibiusky, J., Eisbruch, G. and Simonassi, D. (2012), *Getting started with storm*, O'Reilly Media, Inc.
- [137] Li, H. and Dong, F. (2016), "Research on the implementation strategy of the smart library services", *Library*, Vol. 260, No. 5, pp. 80-84.

- [138] Li, L. (2013), "Designing and implementation of university library automatic management system based on the Internet of Things", In *Joint International Conference on Pervasive Computing and the Networked World*, pp. 241-247, Springer, Cham.
- [139] Li, S., Jiao, F., Zhang, Y. and Xu, X. (2019), "Problems and changes in digital libraries in the age of big data from the perspective of user services", *The Journal of Academic Librarianship*, Vol. 45, No. 1, pp. 22-30.
- [140] Lin, J. and Dyer, C. (2010), "Data-intensive text processing with MapReduce", *Synthesis Lectures on Human Language Technologies*, Vol. 3, No. 1, pp. 1-177.
- [141] Lin, Z. (2014), "An empirical investigation of user and system recommendations in e-commerce", *Decision Support Systems*, Vol. 68, pp. 111-124.
- [142] Linden, G., Smith, B. and York, J. (2003), "Amazon.com recommendations: item-to-item collaborative filtering", *IEEE Internet Computing*, Vol. 7, No. 1, pp. 76-80.
- [143] Lipowski, Z.J. (1975), "Sensory and information inputs overload: Behavioral effects", *Comprehensive Psychiatry*, Vol. 16, No. 3, pp. 199-221.
- [144] Liu, F., Tan, C. W., Lim, E. T. and Choi, B. (2017), "Traversing knowledge networks: an algorithmic historiography of extant literature on the Internet of Things (IoT)", *Journal of Management Analytics*, Vol. 4, No. 1, pp. 3-34.
- [145] Liu, Q. (2015), "Practical exploration on intelligent library in Taipei city", *Library Development*, Vol. 255, No. 9, pp. 56-59.
- [146] Liu, S. and Shen, X. L. (2018), "Library management and innovation in the Big Data Era", *Library Hi Tech*, Vol. 36, No. 3, pp. 374-377.
- [147] Liu, S., Wang, L. and Huang, W. W. (2017), "Effects of process and outcome controls on business process outsourcing performance: Moderating roles of vendor and client capability risks", *European Journal of Operational Research*, Vol. 260, No. 3, pp. 1115-1128.
- [148] Liu, X. and Sheng, W. (2011), "Application on Internet of Things Technology Using in Library Management", In *Advanced Research on Electronic Commerce, Web Application, and Communication*, pp. 391-395, Springer, Berlin, Heidelberg.
- [149] Lops, P., De Gemmis, M. and Semeraro, G. (2011), "Content-based recommender systems: State of the art and trends", In *Recommender systems handbook*, pp. 73-105, Springer, Boston, MA.
- [150] Lu, J. (2004), "A personalized e-learning material recommender system", In *International Conference on Information Technology and Applications*, Macquarie Scientific Publishing.
- [151] Lu, J., Wu, D., Mao, M., Wang, W. and Zhang, G. (2015), "Recommender system application developments: a survey", *Decision Support Systems*, Vol. 74, pp. 12-32.

- [152] Lu, S., Li, R. M., Tjhi, W. C., Lee, K. K., Wang, L., Li, X. and Ma, D. (2011), "A framework for cloud-based large-scale data analytics and visualization: Case study on multiscale climate data", In *Cloud Computing Technology and Science (CloudCom), 2011 IEEE Third International Conference*, pp. 618-622, IEEE.
- [153] Ma, L. (2015), "Digital library network based on the internet of things", *Electronic Engineering and Information Science: Proceedings of the International Conference of Electronic Engineering and Information Science 2015 (ICEEIS 2015)*, CRC Press, Harbin, January 17-18, pp. 231.
- [154] Ma, L., Li, H. and Zhou, J. (2011), "Research and design of study room management system in university library based on Internet of Things", In *Electrical and Control Engineering (ICECE), 2011 International Conference on*, pp. 4033-4035, IEEE.
- [155] Markus, M. L. (2015), "New games, new rules, new scoreboards: the potential consequences of Big Data", *Journal of Information Technology*, Vol. 30, No. 1, pp. 58-59.
- [156] Masthoff, J. (2011), "Group recommender systems: Combining individual models", In *Recommender systems handbook*, pp. 677-702, Springer, Boston, MA.
- [157] McCarthy, K., McGinty, L., Smyth, B. and Salamó, M. (2006), "The needs of the many: a case-based group recommender system", In *European Conference on Case-Based Reasoning*, pp. 196-210, Springer, Berlin, Heidelberg.
- [158] Meier, R.L. (1963), "Communications overload: Proposals from the study of a university library", *Administrative Science Quarterly*, pp. 521-544.
- [159] Melville, P., Mooney, R. J. and Nagarajan, R. (2002), "Content-boosted collaborative filtering for improved recommendations", *Aaai/iaai*, Vol. 23, pp. 187-192.
- [160] Menascé, D. A., Almeida, V. A., Fonseca, R. and Mendes, M. A. (1999), "A methodology for workload characterization of e-commerce sites", In *Proceedings of the 1st ACM conference on Electronic commerce*, pp. 119-128.
- [161] Miller, M. C., Reus, J. F., Matzke, R. P., Koziol, Q. A. and Cheng, A. P. (2004), "Smart libraries: Best SQE practices for libraries with emphasis on scientific computing", In *Proceedings of the Nuclear Explosives Code Developer's Conference*, Vol. 1.
- [162] Milovanović, S., Bogdanović, Z., Labus, A., Barać, D. and Despotović-Zrakić, M. (2019), "An approach to identify user preferences based on social network analysis", *Future Generation Computer Systems*, Vol. 93, pp. 121-129.
- [163] Mooney, R. J. and Roy, L. (2000), "Content-based book recommending using learning for text categorization", In *Proceedings of the fifth ACM conference on Digital libraries*, pp. 195-204.
- [164] Murphy, K. P. (2012), *Machine learning: a probabilistic perspective*, MIT press.

- [165] Nikolaidou, M., Anagnostopoulos, D. and Hatzopoulos, M. (2005), "Development of a medical digital library managing multiple collections", *The Electronic Library*, Vol. 23, No. 2, pp. 221-236.
- [166] Nolin, J. and Olson, N. (2016), "The Internet of Things and convenience", *Internet Research*, Vol. 26, No. 2, pp. 360-376.
- [167] Olson, M. (2010), "Hadoop: Scalable, flexible data storage and analysis", *IQT Quart*, Vol. 1, No. 3, pp. 14-18.
- [168] Owen, S. and Owen, S. (2012), Mahout in action.
- [169] Pal, S. (2016), *SQL on Big Data: Technology, Architecture, and Innovation*. Apress.
- [170] Panetta, K. (2016), *Gartner's Top 10 Strategic Technology Trends for 2017*, доступно на: www.gartner.com (датум приступа: 10.5.2017).
- [171] Pavlo, A., Paulson, E., Rasin, A., Abadi, D. J., DeWitt, D. J., Madden, S. and Stonebraker, M. (2009), "A comparison of approaches to large-scale data analysis", *In Proceedings of the 2009 ACM SIGMOD International Conference on Management of data*, pp. 165-178, ACM.
- [172] Pazzani, M. J. (1999), "A framework for collaborative, content-based and demographic filtering", *Artificial intelligence review*, Vol. 13, No. 5-6, pp. 393-408.
- [173] Philip Chen, C.L. and Zhang, C.Y. (2014), "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data", *Journal of Information Sciences*, Vol. 275, pp. 314-347.
- [174] Pinfield, S., Eaton, J., Edwards, C., Russell, R., Wissenburg, A. and Wynne, P. (1998), "Realising the hybrid library", *New review of information networking*, Vol. 4, No. 1, pp. 3-21.
- [175] Porcel, C. and Herrera-Viedma, E. (2010), "Dealing with incomplete information in a fuzzy linguistic recommender system to disseminate information in university digital libraries", *Knowledge-Based Systems*, Vol. 23, No. 1, pp. 32-39.
- [176] Porcel, C., Moreno, J. M. and Herrera-Viedma, E. (2009), "A multi-disciplinar recommender system to advice research resources in university digital libraries", *Expert Systems with Applications*, Vol. 36, No. 10, pp. 12520-12528.
- [177] Radenković, B., Despotović-Zrakić, M., Bogdanović, Z., Barać, D., Labus, A. and Bojović, Ž. (2017), *Internet inteligentnih uređaja*, Fakultet organizacionih nauka, ISBN:978-86-7680-304-0.
- [178] Raghupathi, W. and Raghupathi, V. (2014), "Big Data analytics in healthcare: promise and potential", *Health information science and systems*, Vol. 2, No. 1, pp. 3.
- [179] Ramakrishnan, R., Sridharan, B., Douceur, J. R., Kasturi, P., Krishnamachari-Sampath, B., Krishnamoorthy, K., ... and Sharman, N. (2017), "Azure data lake store: a hyperscale

- distributed file service for big data analytics", In *Proceedings of the 2017 ACM International Conference on Management of Data*, pp. 51-63, ACM.
- [180] Ranganathan, S. R. (1931), *The five laws of library science*, Madras Library Association, Madras, India, and Edward Goldston, London, UK.
- [181] Ranjan, R., Zhao, L., Wu, X., Liu, A., Quiroz, A. and Parashar, M. (2010), "Peer-to-peer cloud provisioning: Service discovery and load-balancing", In *Cloud Computing*, pp. 195-217, Springer, London.
- [182] Rashid, Z., Melià-Seguí, J., Pous, R. and Peig, E. (2017), "Using Augmented Reality and Internet of Things to improve accessibility of people with motor disabilities in the context of Smart Cities", *Future Generation Computer Systems*, Vol. 76, pp. 248-261.
- [183] Ratledge, D. and Sproles, C. (2017), "An analysis of the changing role of systems librarians", *Library Hi Tech*, Vol. 35, No. 2, pp. 303-311.
- [184] Renaud, J., Britton, S., Wang, D. and Ogihara, M. (2015), "Mining library and university data to understand library use patterns", *The Electronic Library*, Vol. 33, No. 3, pp. 355-372.
- [185] Renda, M. E. and Straccia, U. (2005), "A personalized collaborative digital library environment: a model and an application", *Information processing & management*, Vol. 41, No. 1, pp. 5-21.
- [186] Ricci, F., Rokach, L., Shapira, B., Kantor, PB. (2011), *Recommender Systems Handbook*, Springer, New York.
- [187] Rodrigues, R. A., Lima Filho, L. A., Gonçalves, G. S., Mialaret, L. F., da Cunha, A. M. and Dias, L. A. V. (2018), "Integrating NoSQL, Relational Database, and the Hadoop Ecosystem in an Interdisciplinary Project involving Big Data and Credit Card Transactions", In *Information Technology-New Generations*, Springer, Cham, pp. 443-451.
- [188] Romero, C., Ventura, S., Zafra, A. and De Bra, P. (2009), "Applying Web usage mining for personalizing hyperlinks in Web-based adaptive educational systems", *Computers & Education*, Vol. 53, No. 3, pp. 828-840.
- [189] Rubin, R. E. (2017), *Foundations of library and information science*, American Library Association.
- [190] Rusbridge, C. (1998), *Towards the hybrid library*.
- [191] Saboo, A. R., Kumar, V. and Park, I. (2016), "Using Big Data to Model Time-Varying Effects for Marketing Resource (Re) Allocation", *MIS Quarterly*, Vol. 40, No. 4, pp. 911-939.
- [192] Sagiroglu, S. and Sinanc, D. (2013, May), "Big data: A review", In *Collaboration Technologies and Systems (CTS), 2013 International Conference*, IEEE, pp. 42-47.

- [193] Sakr, S. (2016), *Big Data 2.0 Processing Systems: A Survey*, Springer International Publishing.
- [194] Salton, G., Wong, A. and Yang, C. S. (1975), "A vector space model for automatic indexing", *Communications of the ACM*, Vol. 18, No. 11, pp. 613-620.
- [195] Santos, O. C., Boticario, J. G. and Pérez-Marín, D. (2014), "Extending web-based educational systems with personalised support through User Centred Designed recommendations along the e-learning life cycle", *Science of Computer Programming*, Vol. 88, pp. 92-109.
- [196] Sarwar, B., Karypis, G., Konstan, J. and Riedl, J. (2000), *Application of dimensionality reduction in recommender system-a case study* (No. TR-00-043), Minnesota Univ Minneapolis Dept of Computer Science.
- [197] Sarwar, B., Karypis, G., Konstan, J. and Riedl, J. (2001), "Item-based collaborative filtering recommendation algorithms", In *Proceedings of the 10th international conference on World Wide Web*, pp. 285-295, ACM.
- [198] Schafer, J. B., Konstan, J. A. and Riedl, J. (2001), "E-commerce recommendation applications", *Data mining and knowledge discovery*, Vol. 5, No. 1-2, pp. 115-153.
- [199] Schafer, J. B., Konstan, J. and Riedl, J. (1999), "Recommender systems in e-commerce", In *Proceedings of the 1st ACM conference on Electronic commerce*, pp. 158-166, ACM.
- [200] Schein, A. I., Popescul, A., Ungar, L. H. and Pennock, D. M. (2002), "Methods and metrics for cold-start recommendations", In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 253-260, ACM.
- [201] Serrano-Guerrero, J., Herrera-Viedma, E., Olivas, J. A., Cerezo, A. and Romero, F. P. (2011), "A google wave-based fuzzy recommender system to disseminate information in University Digital Libraries 2.0", *Information Sciences*, Vol. 181, No 9, pp. 1503-1516.
- [202] Shani, G. and Gunawardana, A. (2011), "Evaluating recommendation systems", In *Recommender systems handbook*, pp. 257-297, Springer, Boston, MA.
- [203] Shen, XL., Sun, Y. and Wang, N. (2013), "Recommendations from friends anytime and anywhere: Toward a model of contextual offer and consumption values", *Cyberpsychology, Behavior, and Social Networking*, Vol. 16, No. 5, pp. 349-356.
- [204] Shiwei, W. (2012), "On Three Main Features of the Smart Library [J]", *Journal of Library Science in China*, Vol. 6, pp. 22-28.
- [205] Simović A. (2017), "Big Data Analytics in Public Administration", *XLIV Symposium on Operational Research SYM-OP-IS 2017*, Vol. 44, pp. 382-387, Zlatibor, Srbija.

- [206] Simović A. and Ćirović Z. (2016), "Primena Hadoop okvira u analizi velikih skupova podataka", *XV međunarodni naučno-stručni Simpozijum INFOTEH*, Vol. 15, pp. 593-598, Jahorina.
- [207] Simović, A. (2014), "Sistemi preporuke u e-trgovini", *Sinteza 2014-Impact of the Internet on Business Activities in Serbia and Worldwide*, pp. 846-852, Belgrade, Serbia.
- [208] Simović, A. (2017), "Recommender systems in the big data environment using Mahout framework", In *2017 25th Telecommunication Forum (TELFOR)*, pp. 1-4, IEEE.
- [209] Simović, A. (2018), "A Big Data smart library recommender system for an educational institution", *Library Hi Tech*, Vol. 36, No. 3, pp. 498-523.
- [210] Simović, A. and Ćirović, Z. (2016), "Sistem preporuke u elektronskoj trgovini u Big Data okruženju", *ETRAN 2016*, Vol. 60, pp. RT4.5.1-4, Zlatibor, Srbija.
- [211] Sinclair, B. (2009), "The blended librarian in the learning commons: New skills for the blended library", *College & Research Libraries News*, Vol. 70, No. 9, pp. 504-516.
- [212] Smith, A. (2016), "Big Data Technology, Evolving Knowledge Skills and Emerging Roles", *Legal Information Management*, Vol. 16, No. 4, pp. 219-224.
- [213] Spangler, W. S., Kreulen, J. T., Chen, Y., Proctor, L., Alba, A., Lelescu, A. and Behal, A. (2010), "A smarter process for sensing the information space", *IBM Journal of Research and Development*, Vol. 54, No. 4, pp. 1-13.
- [214] Srinivasan, V. (2016), *The Intelligent Enterprise in the Era of Big Data*, John Wiley & Sons.
- [215] Sun, C. L., Shi, D. and Guo, D. M. (2014), "Analysis on the library digitization construction under the big data era of university", In *Applied Mechanics and Materials*, Vol. 687, pp. 2656-2659.
- [216] Sun, M. (2014), "The Research on the Development of Smart Library", In *Applied Mechanics and Materials*, Vol. 571, pp. 1184-1188.
- [217] Sun, N., Morris, J. G., Xu, J., Zhu, X. and Xie, M. (2014), "iCARE: A framework for Big Data-based banking customer analytics", *IBM Journal of Research and Development*, Vol. 58, No. 5/6, pp. 4-1.
- [218] Svensson, L. G. and Jahns, Y. (2010), "PDF, CSS, RSS and other Acronyms: Redefining the Bibliographic Services of the German National Library", In *World Library and Information Congress: 76th IFLA General Conference and Assembly*, pp. 10-15.
- [219] Teets, M. and Goldner, M. (2013), "Libraries' role in curating and exposing Big Data", *Future internet*, Vol. 5, No. 3, pp. 429-438.
- [220] Tejada, Z. (2017), *Mastering azure analytics: architecting in the cloud with azure data lake, HDInsight, and Spark*, O'Reilly Media, Inc.

- [221] Thaduri, A., Galar, D. and Kumar, U. (2015), "Railway assets: a potential domain for Big Data analytics", *Procedia Computer Science*, Vol. 53, pp. 457-467.
- [222] *The Microsoft Modern Data Warehouse, White paper*, Microsoft Corporation, 2016, доступно на: www.microsoft.com (датум приступа: 12.2.2017).
- [223] Thompson, K. and Works, H. I. (2012), "Underground robot library", *Popular Science*, pp. 56-57.
- [224] Thusoo, A., Sarma, J.S., Jain, N., Shao, Z., Chakka, P., Anthony, S., Liu, H., Wyckoff, P. and Murthy, R. (2009), "Hive: a warehousing solution over a mapreduce framework", *Journal Proceedings of the VLDB Endowment*, Vol. 2, No. 2, pp. 1626-1629.
- [225] Toffler, A. (1971). *Future shock*. Bantam.
- [226] Tsai, C. S. and Chen, M. Y. (2008), "Using adaptive resonance theory and data-mining techniques for materials recommendation based on the e-library environment", *The Electronic Library*, Vol. 26, No. 3, pp. 287-302.
- [227] Tu, Z. (2005), "Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering", In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, Vol. 2, pp. 1589-1596, IEEE.
- [228] Usama, M., Liu, M. and Chen, M. (2017), "Job schedulers for big data processing in hadoop environment: testing real-life schedulers using benchmark programs", *Digital Communications and Networks*, Vol. 3, No. 4, pp. 260-273.
- [229] Vavilapalli, V. K., Murthy, A. C., Douglas, C., Agarwal, S., Konar, M., Evans, R., ... and Saha, B. (2013), "Apache hadoop yarn: Yet another resource negotiator", In *Proceedings of the 4th annual Symposium on Cloud Computing*, ACM.
- [230] Vollmann, T.E. (1991), "Cutting the Gordian knot of misguided performance measurement", *Industrial Management & Data Systems*, Vol. 91, No. 1, pp. 24-26.
- [231] Wadkar, S. and Siddalingaiah, M. (2014), "Apache ambari", In *Pro Apache Hadoop*, pp. 399-401, Apress, Berkeley, CA.
- [232] Wahono, R. S. (2000), "Intelligent agent architecture for digital library", In *Proceedings of the 9th scientific meeting*, pp. 263-266.
- [233] Wang, C., Xu, S., Chen, L. and Chen, X. (2016), "Exposing library data with big data technology: A review", In *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, pp. 1-6, IEEE.
- [234] Wang, L., Da Xu, L., Bi, Z. and Xu, Y. (2014), "Data cleaning for RFID and WSN integration", *IEEE Trans. Industrial Informatics*, Vol. 10, No. 1, pp. 408-418.

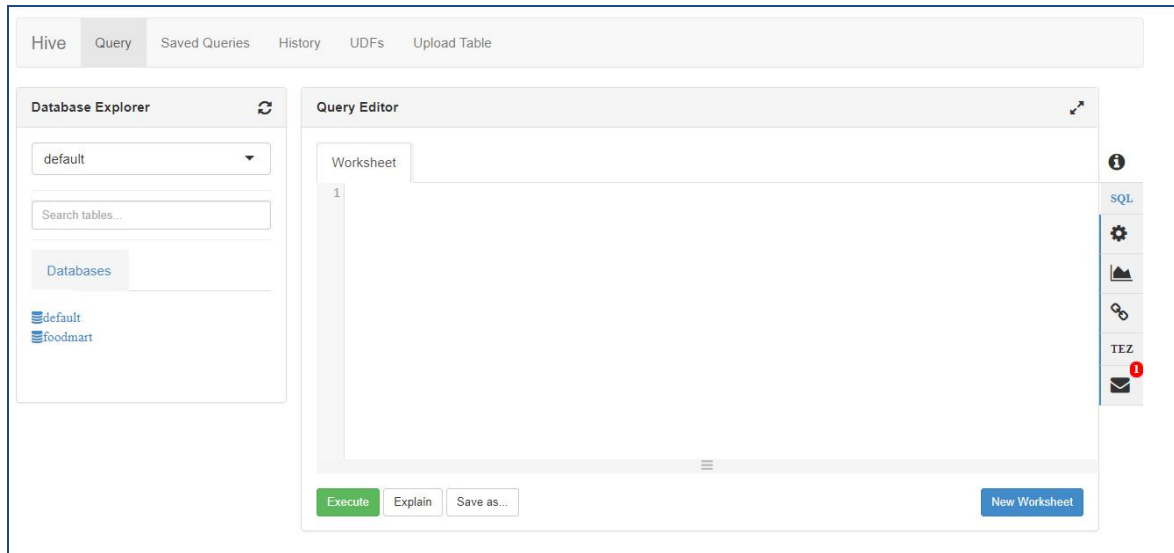
- [235] Wang, L., Tao, J., Ranjan, R., Marten, H., Streit, A., Chen, J. and Chen, D. (2013), "G-Hadoop: MapReduce across distributed data centers for data-intensive computing", *Future Generation Computer Systems*, Vol. 29, No. 3, pp. 739-750.
- [236] Wang, Q. (2016, August), "Design and implementation of recommender system based on Hadoop", In *Software Engineering and Service Science (ICSESS), 2016 7th IEEE International Conference*, IEEE , pp. 295-299.
- [237] Wang, S. (2011), "New pattern of future libraries: the smart library", *Library Development*, Vol. 12, pp. 1-5.
- [238] Wang, S. W. (2017), "Information civilization and library development trend", *Journal of Library Science in China*, Vol. 43, No. 5, pp. 4-20.
- [239] Wang, X., Li, J., Yang, M., Chen, Y. and Xu, X. (2018), "An empirical study on the factors influencing mobile library usage in IoT era", *Library Hi Tech*, Vol. 36, No. 4, pp. 605-621.
- [240] Wang, X.F. (2014), "The application of the internet of things in intelligent digital campus", *Applied Mechanics and Materials*, Vol. 543, pp. 3503-3506.
- [241] Wei, Q. and Yang, Y. (2017), "WeChat Library: a new mode of mobile library service", *The Electronic Library*, Vol. 35, No. 1, pp. 198-208.
- [242] Wei, X. R. L. (2012), "SoLoMo and Smart Libraries [J]", *Journal of Academic Libraries*, Vol. 3, pp. 2.
- [243] Wei, Y., Sha, F. and Yan, W. (2014), "The Construction of Information Management System Based on Cloud Computing and the Internet of Things", *Applied Mechanics & Materials*, Vol. 543, pp. 2981-2983, Trans Tech Publications Ltd.
- [244] White, T. (2012), *Hadoop: The definitive guide*, O'Reilly Media, Inc.
- [245] Wójcik, M. (2016), "Internet of Things—potential for libraries", *Library Hi Tech*, Vol. 34, No. 2, pp. 404-420.
- [246] Wooldridge, M. and Jennings, N. R. (1995), "Intelligent agents: Theory and practice", *The knowledge engineering review*, Vol. 10, No. 2, pp. 115-152.
- [247] Wu, L., Shah, S., Choi, S., Tiwari, M. and Posse, C. (2014), "The Browsemaps: collaborative filtering at LinkedIn", In *RSWeb@ RecSys*.
- [248] Wurman, R. S. (2001), *Information anxiety*, No. 302.234, WUR. CIMMYT.
- [249] Xianchun, L. and Ming, Y. (2014), "Evolution of public library value: According to research perspective changes of value system", *Journal of Library Science in China*, Vol. 40, No. 3, pp. 27-36.
- [250] Xiaolin, Z. (2011), "The Trends That Will Disrupt Digital Libraries [J] ", *Journal of Library Science in China*, Vol. 5, pp. 4-12.

- [251] Xiaoxia, D., Xiangyang, G., Ruolin, Z. and Chaobin, Y. (2011), "The Design and Implementation of Smart Library", *Data Analysis and Knowledge Discovery*, Vol. 27, No. 2, pp. 76-80.
- [252] Xinling, S. U. (2015), "Opportunities and challenges faced by digital libraries in the era of big data", *Journal of Library Science in China*, Vol. 41, No. 6, pp. 4-12.
- [253] Xu, B., Da Xu, L., Cai, H., Xie, C., Hu, J. and Bu, F. (2014), "Ubiquitous data accessing method in IoT-based information system for emergency medical services", *IEEE Trans. Industrial Informatics*, Vol. 10, No. 2, pp. 1578-1586.
- [254] Xu, L. (2014), "The Internet of Things technology application and the intelligent library", In *Applied Mechanics and Materials*, Vol. 571, pp. 1180-1183.
- [255] Xu, L. D., Xu, E. L. and Li, L. (2018), "Industry 4.0: state of the art and future trends", *International Journal of Production Research*, Vol. 56, No. 8, pp. 2941-2962.
- [256] Xu, S., Du, W., Wang, C. and Liu, D. (2017), "The library big data research: status and directions", *International Journal of Software Innovation (IJSI)*, Vol. 5, No. 3, pp. 77-88.
- [257] Yan, D. (2010), "Smart library based on the Internet of Things", *Journal of Library Science*, Vol. 32, No. 7, pp. 8-10.
- [258] Yao, G. Y. and Song, L. (2014), "Design of library lighting energy saving system based on Internet of Things", In *Applied Mechanics and Materials*, Vol. 496, pp. 1690-1693, Trans Tech Publications.
- [259] Yao, Q., Tian, Y., Li, P. F., Tian, L. L., Qian, Y. M. and Li, J. S. (2015), "Design and development of a medical big data processing system based on Hadoop", *Journal of medical systems*, Vol. 39, No. 3, pp. 23.
- [260] Yi, K., Chen, T. and Cong, G. (2018), "Library personalized recommendation service method based on improved association rules", *Library Hi Tech*, Vol. 36, No. 3, pp. 443-457.
- [261] Yu, T., Chawla, N. and Simoff, S. J. (2013), *Computational intelligent data analysis for sustainable development*, CRC Press.
- [262] Zaiane, O. R. (2002), "Building a recommender agent for e-learning systems", In *Computers in education, 2002. proceedings. international conference*, IEEE, pp. 55-59.
- [263] Zhao, X. (2013), "The Research and Construction of Wisdom Campus Based on the Internet of Things", In *Intelligent Networks and Intelligent Systems (ICINIS), 2013 6th International Conference on*, pp. 162-166, IEEE.
- [264] Zikopoulos, P. and Eaton, C. (2011), *Understanding big data: Analytics for enterprise class hadoop and streaming data*, McGraw-Hill Osborne Media.

- [265] Zimmerman, T. and Chang, H. C. (2018), "Getting Smarter: Definition, Scope, and Implications of Smart Libraries", *In Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*, ACM, pp. 403-404.
- [266] Zomaya, A. Y. and Sakr, S. (2017), *Handbook of big data technologies*, Berlin: Springer.
- [267] Zong, C. J., Jia, B. X. and Zhang, Y. (2014), "Research on Application of the Internet of Things in University's Teaching Management", *In Advanced Materials Research*, Vol. 860, pp. 3017-3020.

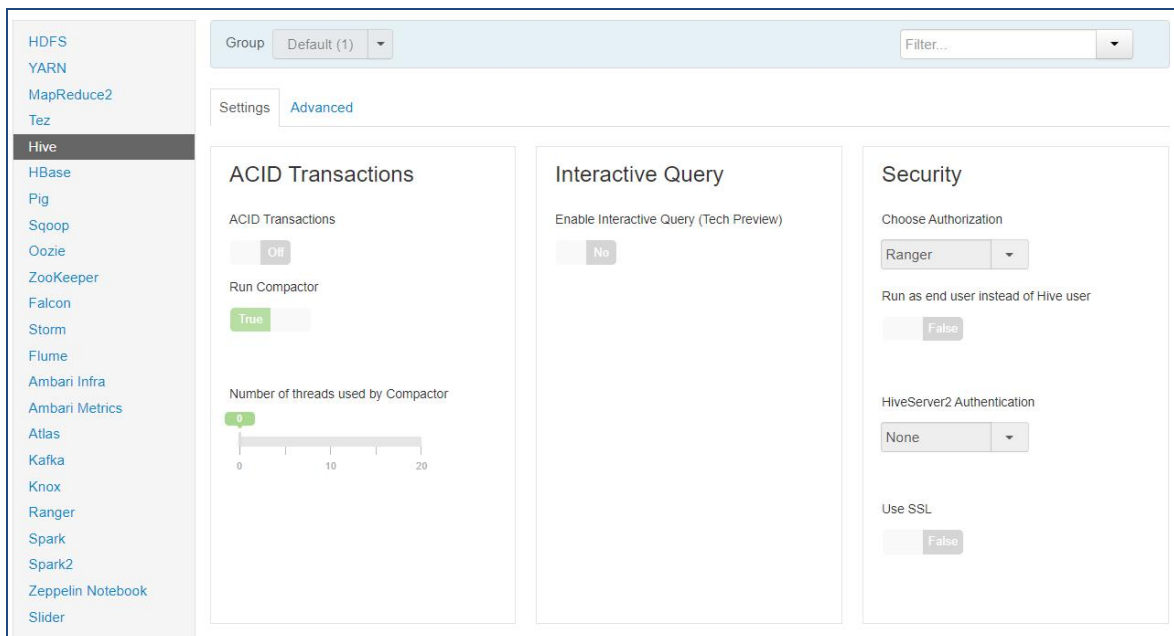
11. ПРИЛОЗИ

ПРИЛОГ А1



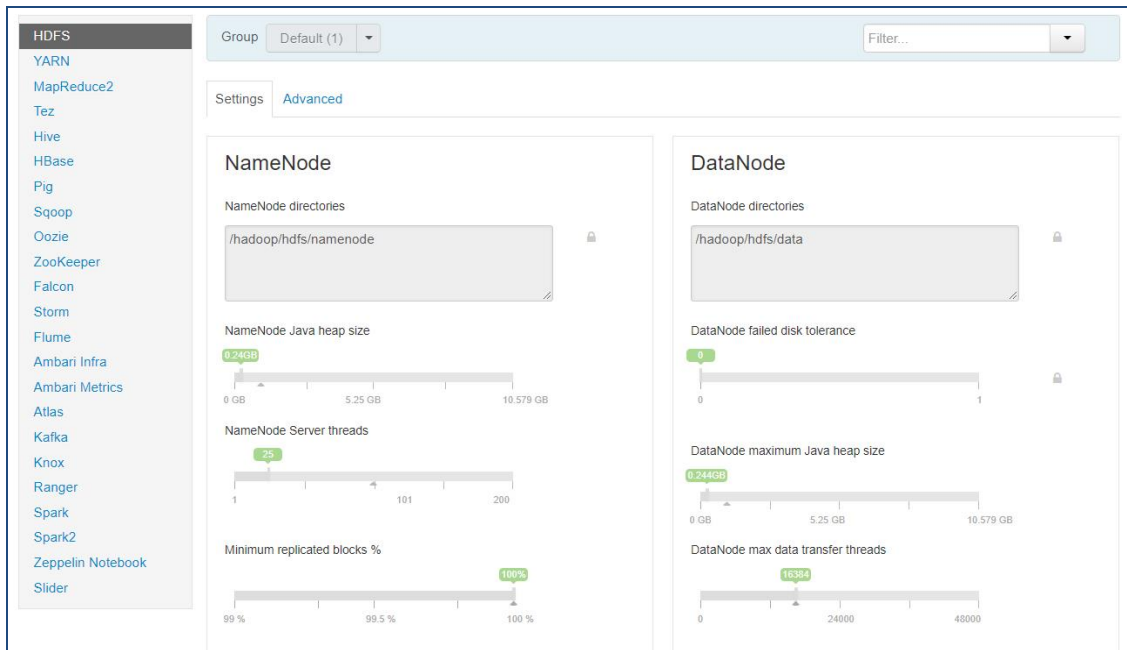
Опис: Подразумевани приступ Hive сервису из Ambari модула

ПРИЛОГ А2



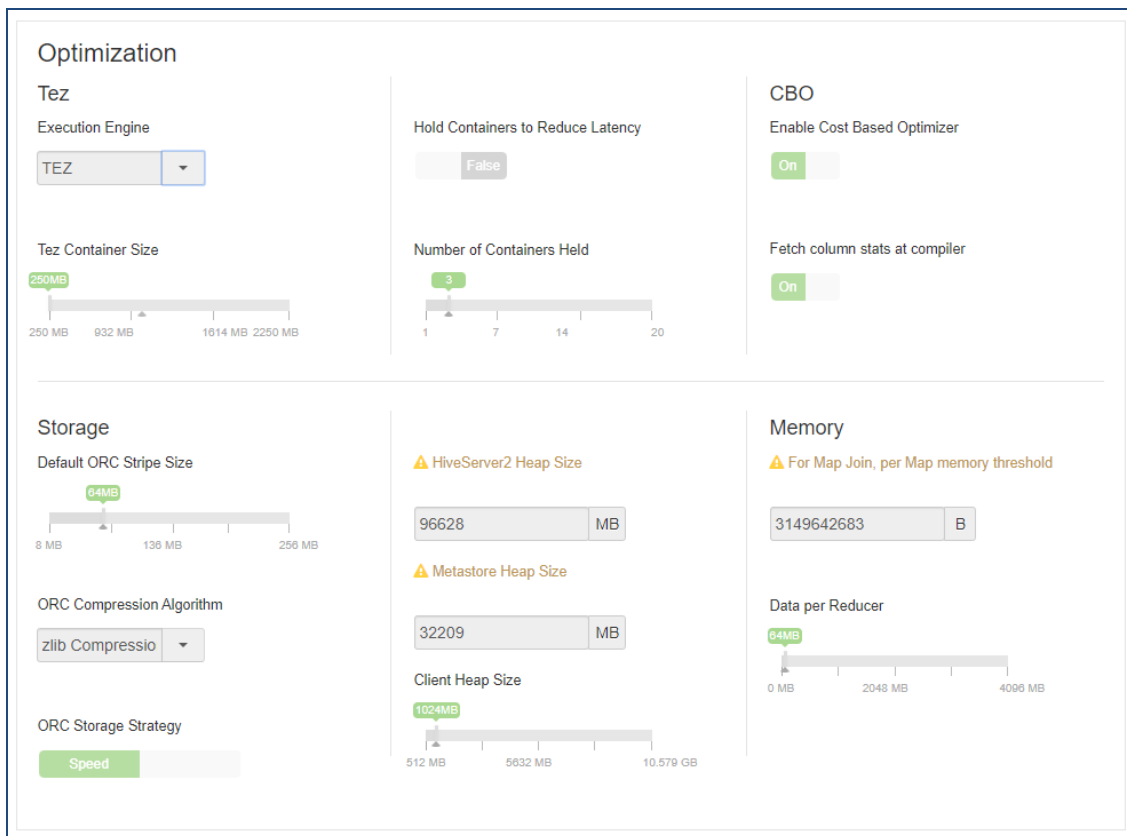
Опис: Конфигурација Hive сервиса

ПРИЛОГ А3



Опис: Hadoop дистрибуирани фајл систем

ПРИЛОГ А4



Опис: Tez оптимизација

ПРИЛОГ А5

Service	Version	Status	Description
HDFS	2.7.1.2.5	Installed	Apache Hadoop Distributed File System
YARN	2.7.1.2.5	Installed	Apache Hadoop NextGen MapReduce (YARN)
MapReduce2	2.7.1.2.4	Installed	Apache Hadoop NextGen MapReduce (YARN)
Tez	0.7.0.2.5	Installed	Tez is the next generation Hadoop Query Processing framework written on top of YARN.
Hive	1.2.1.2.5	Installed	Data warehouse system for ad-hoc queries & analysis of large datasets and table & storage management service
HBase	1.1.2.2.5	Installed	A Non-relational distributed database, plus Phoenix, a high performance SQL layer for low latency applications.
Pig	0.16.0.2.5	Installed	Scripting platform for analyzing large datasets
Sqoop	1.4.6.2.5	Installed	Tool for transferring bulk data between Apache Hadoop and structured data stores such as relational databases
Oozie	4.2.0.2.5	Installed	System for workflow coordination and execution of Apache Hadoop jobs. This also includes the installation of the optional Oozie Web Console which relies on and will install the ExtJS Library.
ZooKeeper	3.4.6.2.5	Installed	Centralized service which provides highly reliable distributed coordination
Falcon	0.10.0.2.5	Installed	Data management and processing platform
Storm	1.0.1.2.5	Installed	Apache Hadoop Stream processing framework
Flume	1.5.2.2.5	Installed	A distributed service for collecting, aggregating, and moving large amounts of streaming data into HDFS
Accumulo	1.7.0.2.5	Add Service	Robust, scalable, high performance distributed key/value store.
Ambari Infra	0.1.0	Installed	Core shared service used by Ambari managed components.
Ambari Metrics	0.1.0	Installed	A system for metrics collection that provides storage and retrieval capability for metrics collected from the cluster
Atlas	0.7.0.2.5	Installed	Atlas Metadata and Governance platform

Опис: Покренути и конфигурисани софтверски сервиси Hadoop екосистема модела паметне библиотеке

ПРИЛОГ А6

Kafka	0.10.0.2.5	Installed	A high-throughput distributed messaging system
Knox	0.9.0.2.5	Installed	Provides a single point of authentication and access for Apache Hadoop services in a cluster
Log Search	0.5.0	Add Service	Log aggregation, analysis, and visualization for Ambari managed services. This service is Technical Preview .
Ranger	0.6.0.2.5	Installed	Comprehensive security for Hadoop
Ranger KMS	0.6.0.2.5	Add Service	Key Management Server
SmartSense	1.3.0.0-1225	Add Service	SmartSense - Hortonworks SmartSense Tool (HST) helps quickly gather configuration, metrics, logs from common HDP services that aids to quickly troubleshoot support cases and receive cluster-specific recommendations.
Spark	1.6.x.2.5	Installed	Apache Spark is a fast and general engine for large-scale data processing.
Spark2	2.0.x.2.5	Installed	Apache Spark 2.0 is a fast and general engine for large-scale data processing. This service is Technical Preview .
Zeppelin Notebook	0.6.0.2.5	Installed	A web-based notebook that enables interactive data analytics. It enables you to make beautiful data-driven, interactive and collaborative documents with SQL, Scala and more.
Kerberos	1.10.3-10	Add Service	A computer network authentication protocol which works on the basis of 'tickets' to allow nodes communicating over a non-secure network to prove their identity to one another in a secure manner.
Mahout	0.9.0.2.5	Add Service	Project of the Apache Software Foundation to produce free implementations of distributed or otherwise scalable machine learning algorithms focused primarily in the areas of collaborative filtering, clustering and classification
NiFi	1.0.0-DEMO	Add Service	Apache NiFi is an easy to use, powerful, and reliable system to process and distribute data. This service is for demo purposes only and not officially supported
Slider	0.80.0.2.5	Installed	A framework for deploying, managing and monitoring existing distributed applications on YARN.
Solr	5.5.2.2.5	Add Service	Solr is a search platform from the Apache Lucene project. Its major features include full-text search, hit highlighting, faceted search, dynamic clustering, database integration, and rich document (e.g., Word, PDF) handling.

Опис: Покренути и конфигурирани софтверски сервиси Hadoop екосистема модела паметне библиотеке – наставак прилога А5

ПРИЛОГ А7

HDFS	2.7.1.2.5	Storm	1.0.1.2.5
YARN	2.7.1.2.5	Flume	1.5.2.2.5
MapReduce2	2.7.1.2.4	Ambari Infra	0.1.0
Tez	0.7.0.2.5	Ambari Metrics	0.1.0
Hive	1.2.1.2.5	Atlas	0.7.0.2.5
HBase	1.1.2.2.5	Kafka	0.10.0.2.5
Pig	0.16.0.2.5	Knox	0.9.0.2.5
Sqoop	1.4.6.2.5	Ranger	0.6.0.2.5
Oozie	4.2.0.2.5	Spark	1.6.x.2.5
ZooKeeper	3.4.6.2.5	Spark2	2.0.x.2.5
Falcon	0.10.0.2.5	Zeppelin Notebook	0.6.0.2.5
		Slider	0.80.0.2.5

Опис: Филтрирање верзија инсталираних модула

ПРИЛОГ А8

Oozie Server

Oozie Server host: sandbox.hortonworks.com

Oozie Database: New Derby Database
 Existing MySQL / MariaDB Database
 Existing PostgreSQL Database
 Existing Oracle Database
 Existing SQL Anywhere Database

Database Name: oozie

Database Username: oozie

Database Password:

JDBC Driver Class: org.apache.derby.jdbc.EmbeddedDriver

Database URL: jdbc:derby:\${oozie.data.dir}/\${oozie.db.schema.name}-db;create=true

Oozie Data Dir: /hadoop/oozie/data

Опис: Конфигурација Oozie сервера

ПРИЛОГ А9

oozie.service.ELService. ext.functions.coord- action-create	now=org.apache.oozie.extensions.OozieELExtensions#ph2_now, today=org.apache.oozie.extensions.OozieELExtensions#ph2_today, yesterday=org.apache.oozie.extensions.OozieELExtensions#ph2_yesterday, currentWeek=org.apache.oozie.extensions.OozieELExtensions#ph2_currentWeek, lastWeek=org.apache.oozie.extensions.OozieELExtensions#ph2_lastWeek, currentMonth=org.apache.oozie.extensions.OozieELExtensions#ph2_currentMonth, lastMonth=org.apache.oozie.extensions.OozieELExtensions#ph2_lastMonth,
oozie.service.ELService. ext.functions.coord- action-create-inst	now=org.apache.oozie.extensions.OozieELExtensions#ph2_now_inst, today=org.apache.oozie.extensions.OozieELExtensions#ph2_today_inst, yesterday=org.apache.oozie.extensions.OozieELExtensions#ph2_yesterday_inst, currentWeek=org.apache.oozie.extensions.OozieELExtensions#ph2_currentWeek_inst, lastWeek=org.apache.oozie.extensions.OozieELExtensions#ph2_lastWeek_inst, currentMonth=org.apache.oozie.extensions.OozieELExtensions#ph2_currentMonth_inst, lastMonth=org.apache.oozie.extensions.OozieELExtensions#ph2_lastMonth_inst,
oozie.service.ELService. ext.functions.coord- action-start	now=org.apache.oozie.extensions.OozieELExtensions#ph2_now, today=org.apache.oozie.extensions.OozieELExtensions#ph2_today, yesterday=org.apache.oozie.extensions.OozieELExtensions#ph2_yesterday, currentWeek=org.apache.oozie.extensions.OozieELExtensions#ph2_currentWeek, lastWeek=org.apache.oozie.extensions.OozieELExtensions#ph2_lastWeek, currentMonth=org.apache.oozie.extensions.OozieELExtensions#ph2_currentMonth, lastMonth=org.apache.oozie.extensions.OozieELExtensions#ph2_lastMonth,
oozie.service.ELService. ext.functions.coord-job- submit-data	now=org.apache.oozie.extensions.OozieELExtensions#ph1_now_echo, today=org.apache.oozie.extensions.OozieELExtensions#ph1_today_echo, yesterday=org.apache.oozie.extensions.OozieELExtensions#ph1_yesterday_echo, currentWeek=org.apache.oozie.extensions.OozieELExtensions#ph1_currentWeek_echo, lastWeek=org.apache.oozie.extensions.OozieELExtensions#ph1_lastWeek_echo,

Опис: Конфигурација Oozie сервера – наставак прилога А8

ПРИЛОГ А10

ZooKeeper directory	<input type="text" value="/hadoop/zookeeper"/>
Length of single Tick	<input type="text" value="2000"/> <input type="text" value="ms"/>
Ticks to allow for sync at Init	<input type="text" value="10"/>
Ticks to allow for sync at Runtime	<input type="text" value="5"/>
Port for running ZK Server	<input type="text" value="2181"/>
<hr/>	
▼ Advanced zoo.cfg	
autopurge.purgeInterval	<input type="text" value="24"/>
autopurge.snapRetainCount	<input type="text" value="30"/>
<hr/>	
▼ Advanced zookeeper-env	
ZooKeeper Log Dir	<input type="text" value="/var/log/zookeeper"/>
ZooKeeper PID Dir	<input type="text" value="/var/run/zookeeper"/>
Zookeeper Server Maximum Memory	<input type="text" value="1024"/> <input type="text" value="MB"/>
zookeeper-env template	<pre>export JAVA_HOME={{java64_home}} export ZOOKEEPER_HOME={{zk_home}} export ZOO_LOG_DIR={{zk_log_dir}} export ZOOIDFILE={{zk_pid_file}} export SERVER_JVMFLAGS={{zk_server_heapsize}} export JAVA=\$JAVA_HOME/bin/java export CLASSPATH=\$CLASSPATH:/usr/share/zookeeper/* {% if security_enabled %} export SERVER_JVMFLAGS="\$SERVER_JVMFLAGS -Djava.security.auth.login.config={{zk_server_jaas_file}}" export CLIENT_JVMFLAGS="\$CLIENT_JVMFLAGS -Djava.security.auth.login.config={{zk_client_jaas_file}}" {% endif %}</pre>

Опис: Конфигурација ZooKeeper модула

ПРИЛОГ Б1

The screenshot displays the Ambari web interface for Hive. The top navigation bar includes 'Ambari', 'Sandbox', and status indicators for '0 ops' and '0 alerts'. The main navigation includes 'Dashboard', 'Services', 'Hosts', 'Alerts', and 'Admin'. The 'Hive' section is active, showing 'Query', 'Saved Queries', 'History', and 'UDFs' tabs. The 'Database Explorer' on the left shows a tree view of databases, with 'smart_library' selected. The 'Query Editor' contains the following SQL query:

```
1 SELECT recommender.book_id,book.name,
2 count(recommender.book_id) OVER (PARTITION BY recommender.person_id ORDER BY recommender.book_id
3 ROWS BETWEEN UNBOUNDED PRECEDING AND UNBOUNDED FOLLOWING)
4 FROM recommender join book on (recommender.book_id=book.count);
```

Buttons for 'Execute', 'Explain', 'Save as...', and 'New Worksheet' are visible. The 'Query Process Results (Status: Succeeded)' section shows a table of results:

recommender.book_id	book.name	count_window_0
1016	Osnovi elektronike i telekomunikacija	3
844	Хидраулика: увод са примерима управљања	2
1002	Моторна возила I: општи и теоријски део	2
138	Електричне машине: за трећи разред електротехничке школе	3
1463	Комутаторни мотори	3
1497	Термоелектране	3
103	Физика: механика чврстих	6
118	Трансформатори	6
571	Психолошки услови трансфера учења	6
1916	Енергетски трансформатори и генератори	6
2113	Фотонапонска постројења: планирање	6
2275	Соларне технологије: тоplotни и фотоелектрични системи	6
7	Основи рачунарске технике	2
264	Математички приручник	2

Опис: Ambari сервис и резултати извршеног упита како би се одредиле књиге које су највише препоручиване студентима применом система препоруке паметне библиотеке засноване на big data технологијама

ПРИЛОГ Б2

The screenshot displays the Ambari web interface. On the left is the Database Explorer showing a tree view of databases, with 'smart_library' selected. The main area is the Query Editor, containing a Hive SQL query. Below the editor are buttons for 'Execute', 'Explain', and 'Save as...'. The 'Query Process Results' section shows a table of results with columns 'recommender.book_id', 'book.name', and 'count_window_0'.

```

2016_borrowing
1 SELECT recommender.book_id, book.name,
2 COUNT(recommender.book_id) OVER (PARTITION BY recommender.book_id ORDER BY recommender.book_id
3 ROWS BETWEEN UNBOUNDED PRECEDING AND UNBOUNDED FOLLOWING)
4 FROM recommender JOIN book ON (recommender.book_id=book.count)
5 WHERE recommender.book_id IN ('1016', '844', '1002', '138', '1463', '1497', '103',
6 '118', '571', '1916', '2113', '2275', '7', '264')
7 AND recommender.date_taken > '2016-1-1' AND recommender.date_return < '2017-1-1';
    
```

recommender.book_id	book.name	count_window_0
7	Основи рачунарске технике	2
7	Основи рачунарске технике	2
103	Физика: механика чврстих	1
118	Трансформатори	1
138	Електричне машине: за трећи разред електротехничке школе	1
264	Математички приручник	1
571	Психолошки услови трансфера учења	1
844	Хидраулика: увод са примерима управљања	1
1002	Моторна возила I: општи и теоријски део	1
1016	Osnovi elektronike i telekomunikacija	1
1463	Komutatorni motori	1
1497	Termoelektrane	2
1497	Termoelektrane	2
1916	Energetski transformatori i generatori	1
2113	Fotonaponska postrojenja: planiranje	1
2275	Solarne tehnologije: toplotni i fotoelektrični sistemi	2
2275	Solarne tehnologije: toplotni i fotoelektrični sistemi	2

Опис: Амбари сервис и резултати извршеног упита. Колико пута су препоручене књиге изнајмљене из библиотеке образовне институције у 2016. години – пре активације big data система препоруке паметне библиотеке

ПРИЛОГ БЗ

The screenshot displays the Ambari web interface for Hive. The top navigation bar includes 'Ambari Sandbox', 'Dashboard', 'Services', 'Hosts', 'Alerts', and 'Admin'. Below this is a 'Hive' section with 'Query', 'Saved Queries', 'History', and 'UDFs' tabs. On the left is a 'Database Explorer' showing a tree view of databases including 'smart_library'. The main area is the 'Query Editor' with a SQL query and an 'Execute' button. Below the editor is the 'Query Process Results' section, which is 'Succeeded' and shows a table of results with columns 'recommender.book_id', 'book.name', and 'count_window_0'.

```

2017_borrowing *
1 SELECT recommender.book_id, book.name,
2 COUNT(recommender.book_id) OVER (PARTITION BY recommender.book_id ORDER BY recommender.book_id
3 ROWS BETWEEN UNBOUNDED PRECEDING AND UNBOUNDED FOLLOWING)
4 FROM recommender JOIN book ON (recommender.book_id=book.count)
5 WHERE recommender.book_id IN ('1016', '844', '1002', '138', '1463', '1497', '103',
6 '118', '571', '1916', '2113', '2275', '7', '264')
7 AND recommender.date_taken>'2017-1-1' AND recommender.date_return<'2017-12-1';
    
```

recommender.book_id	book.name	count_window_0
7	Основи рачунарске технике	2
7	Основи рачунарске технике	2
103	Физика: механика черстих	1
118	Трансформатори	2
118	Трансформатори	2
138	Електричне машине: за трећи разред електротехничке школе	1
264	Математички приручник	1
571	Психолошки услови трансфера учења	1
844	Хидраулика: увод са примерима управљања	2
844	Хидраулика: увод са примерима управљања	2
1002	Моторна возила I: општи и теоријски део	1
1016	Osnovi elektronike i telekomunikacija	2
1016	Osnovi elektronike i telekomunikacija	2
1463	Komutatori motori	1
1497	Termoelektrane	2
1497	Termoelektrane	2
1916	Energetski transformatori i generatori	5
1916	Energetski transformatori i generatori	5
1916	Energetski transformatori i generatori	5
1916	Energetski transformatori i generatori	5
1916	Energetski transformatori i generatori	5
2113	Fotonaponska postrojenja: planiranje	3
2113	Fotonaponska postrojenja: planiranje	3
2113	Fotonaponska postrojenja: planiranje	3
2275	Solarnе tehnologije: toplinski i fotoelektrični sistemi	2

Опис: Амбари сервис и резултати извршеног упита. Колико пута су препоручене књиге изнајмљене из библиотеке образовне институције у 2017. години – током тестирања big data система препоруке паметне библиотеке

11.1. СПИСАК СЛИКА

Слика 1. Комплексност big data концепта у модерном електронском пословању

Слика 2. Инфраструктура дистрибуираног складишта – HDFS

Слика 3. Инфраструктура MapReduce софтверског оквира

Слика 4. MapReduce – Input Split / Shuffle and Sort

Слика 5. Инфраструктура PolyBase технологије

Слика 6. Big data сервиси – Технолошки оквир Hadoop екосистема

Слика 7. Апстрактни приказ Hadoop сервиса

Слика 8. Чворови сервиса Storm кластера

Слика 9. Топологија Storm сервиса у кластеру

Слика 10. Визуализација коришћеног скупа података

Слика 11. Коначно израчунавање и креирање модела евалуације машинског учења

Слика 12. Визуализација проширеног скупа података

Слика 13. Резултати евалуације експеримента big data аналитике

Слика 14. Основни модел преференција корисника

Слика 15. Систем препоруке базиран на садржају

Слика 16. Систем колаборативног филтрирања

Слика 17. Матрица корисника и производа

Слика 18. Модел корисника у персонализованом систему

Слика 19. Неперсонализовани систем препоруке

Слика 20. Реализован систем препоруке са приказом генерисане листе

Слика 21. Колаборативно филтрирање по корисницима – део референтног кода

Слика 22. Генерисана препорука колаборативног филтрирања по корисницима

Слика 23. Колаборативно филтрирање по производима – део референтног кода

Слика 24. Генерисана препорука колаборативног филтрирања по производима

Слика 25. Евалуација система базираног на филтрирању по корисницима – део референтног

кôда за $n/10$ од укупног броја корисника система

Слика 26. Евалуација система базираног на филтрирању по корисницима – део референтног кôда за $n/50$ од укупног броја корисника система

Слика 27. Визуализација добијених резултата израчунавања

Слика 28. Кључне карактеристике модела паметне библиотеке заснованог на *big data* технологијама

Слика 29. Блок дијаграм архитектуре система паметне библиотеке

Слика 30. Компоненте модела паметне библиотеке заснованог на *big data* технологијама – приказ структуре модела са међусобним релацијама

Слика 31. Архитектура предложеног хардверског система паметне библиотеке – *Master / Slave* мултинод кластер – Основна конфигурација система

Слика 32. Детаљан приказ пројектоване *Master / Slave* архитектуре мултинод кластера

Слика 33. Управљање операцијама *Master / Slave* мултинод кластера

Слика 34. Псеудо кôд у задатој *MapReduce* сесии (адаптирано према: *Dean and Ghemawat, 2008*)

Слика 35. *MapReduce Shuffle & Sort* процес

Слика 36. *HDFS* процес пројектованог модела паметне библиотеке

Слика 37. *Big data* технолошки оквир модела паметне библиотеке

Слика 38. Кориснички приступ систему паметне библиотеке

Слика 39. Приступ корисника интегрисаним компонентама система паметне библиотеке са међусобним релацијама интеграције и интеракције података

Слика 40. Кориснички дијаграм (енгл. *Use Case*) приступа апликација система паметне библиотеке – Електронска продавница књига

Слика 41. Кориснички дијаграм (енгл. *Use Case*) приступа апликација система паметне библиотеке – Информациони и библиотечки информациони систем образовне институције

Слика 42. Кориснички дијаграм (енгл. *Use Case*) приступа апликација система паметне библиотеке – *Hadoop* кластер & *LMS Moodle*

Слика 43. Дијаграм тока модела *big data* екосистема паметне библиотеке

Слика 44. *HCatalog* – листа учитаних табела

Слика 45. Метод за агрегацију података и аналитику текста прикупљених са друштвених

медија

Слика 46. Дијаграм тока преноса података о локацијама штампаних издања књига у библиотеци

Слика 47. Приказ резултата израчунавања у *Ambari* сервису паметне библиотеке

Слика 48. Систем препоруке паметне библиотеке заснован на *big data* технологијама, визуализован на *LMS Moodle* платформи

Слика 49. Колико пута су препоручене књиге изнајмљене из библиотеке

11.2. СПИСАК ТАБЕЛА

Табела 1. Дефиниције паметне библиотеке

Табела 2. Категоризација концептуалних релација димензија и дефиниција паметне библиотеке

Табела 3. Идентификациони број корисника и наслови књига

Табела 4. Систем препоруке паметне библиотеке

Табела 5. Књиге које су највише препоручиване студентима путем *LMS Moodle* платформе засноване на *Nadoop* екосистему

Табела 6. Упоредни приказ изнајмљених књига у 2016. у односу на 2017. годину

Табела 7. Онлајн упитник за прикупљање података од студената на основу њихових утисака

Табела 8. Дескриптивна статистика ($n=132$) студената базирана на основу њихових утисака, са приказом средње оцене и стандардним одступањем

11.3. БИОГРАФСКИ ПОДАЦИ О КАНДИДАТУ

Александар Симовић је рођен 5. октобра 1973. године у Београду где завршава основну школу и средњу Електротехничку школу Никола Тесла. Дипломира на Високој школи електротехнике и рачунарства струковних студија у Београду (студијски програм: Електронско пословање) 2011. године са просечном оценом 9,62 и оценом 10 за завршни рад на тему *Израда online продавнице*, а потом и на Техничком факултету Универзитета Сингидунум (студијски програм: Инжењерски менаџмент) 2013. године са просечном оценом 9,76 и оценом 10 за завршни рад на тему *Имплементација интелигентних агената у е-трговини на примеру е-продавнице књига*. Мастер академске студије завршава 2015. године на Факултету организационих наука Универзитета у Београду (студијски програм: Електронско пословање и управљање системима) са просечном оценом 10 и највишом оценом за завршни рад на тему *Системи препоруке у електронској трговини*. Докторске студије уписује на Факултету организационих наука Универзитета у Београду на студијском програму Информациони системи и квантитативни менаџмент на изборном подручју Електронско пословање 2016. године. Положио је свих девет програмом предвиђених испита на докторским студијама са просечном оценом 10. Приступни рад брани 12. септембра 2018. године. Тренутно је запослен као асистент на Високој школи струковних студија за Информационе технологије у Београду.

Каријеру је започео у привредном сектору. Предводио је тим за одржавање дела информационог система ЈАТ-а од 2000. до 2003. године где стиче значајно искуство у домену информационих система и рачунарских мрежа. Пословни напредак наставља у ИТ сектору као шеф сервиса и производње где успешно води тим инжењера. Године 2005. активно почиње са израдом и имплементацијом интерактивних, динамичких веб апликација и *CMS* решења за мале и средње компаније и стиче значајно искуство у домену информационих технологија и електронског пословања. Године 2010. почиње рад у Високој школи електротехнике и рачунарства струковних студија у Београду на предмету Електронска трговина. Исте године израђује електронску продавницу књига за школу, чији је био администратор и модератор до 2018. године. Од 2011. године поверава му се настава и из следећих предмета: Електронско банкарство, Интернет маркетинг, Веб дизајн, Пословни софтвер, Бизнис план за е-пословање, Маркетинг, Интернет сервис, Основи електронског пословања, Интеракција човек-рачунар и Увод у Интернет технологије. Године 2013. изабран је у звање сарадник у настави. Као коаутор је објавио следеће публикације:

[1] *Приручник за вежбе из предмета: Електронска трговина, ISBN: 978-86-7982-116-4, ВИШЕР, Београд, 2012.*

[2] *Приручник за вежбе из предмета: Електронска трговина, ново издање, ISBN: 978-86-7982-194-2, ВИШЕР, Београд, 2014.*

[3] *Електронско банкарство: приручник за лабораторијске вежбе, ISBN: 978-86-7982-202-4, ВИШЕР, Београд, 2014.*

[4] *Електронско банкарство: приручник за лабораторијске вежбе, друго, измењено и*

допуњено издање, ISBN: 978-86-7982-233-8, ВИШЕР, Београд, 2015.

[5] *Основи електронског пословања: приручник*, ISBN: 978-86-7982-239-0, ВИШЕР, Београд, 2016.

[6] *Веб дизајн: приручник за лабораторијске вежбе*, ISBN: 978-86-7982-272-7, ВИШЕР, Београд, 2018.

Године 2018. почиње рад у Високој школи за информационе технологије ИТС у Београду, на предметима: Напредни веб дизајн и Интернет програмерски алати, Програмирање у реалном времену, Веб дизајн и Интернет маркетинг и маркетинг. Исте године је изабран у звање асистента. Године 2019. поверава му се и настава из предмета Основи програмирања, Веб програмирање, Анализа великих података (*Big data*).

Током досадашњег рада у образовању и током школовања, Александар Симовић је објавио више радова у земљи и иностранству и учествовао на више међународних и домаћих скупова и конференција. Резултати истраживања докторске дисертације су објављени у међународном часопису категорије М22.

РАДОВИ ОБЈАВЉЕНИ У ИСТАКНУТИМ МЕЂУНАРОДНИМ ЧАСОПИСИМА (М20):

[1] Simović A. (2018), *A Big Data smart library recommender system for an educational institution*, Library Hi Tech, Vol. 36, Issue: 3, pp. 498-523, doi: 10.1108/LHT-06-2017-0131, ИФ (2018) = 1.256, М22.

РАДОВИ У ЗБОРНИЦИМА МЕЂУНАРОДНИХ СКУПОВА ШТАМПАНИ У ЦЕЛИНИ (М33):

[1] Simović, A. (2017), *Recommender systems in the big data environment using Mahout framework*, In Telecommunication Forum TELFOR 2017, Vol. 25, pp. 1-4, IEEE, doi: 10.1109/TELFOR.2017.8249472, Belgrade, Serbia, ISBN 978-1-5386-3072-3.

[2] Ćirović Z., Simović A., Ćirović N. (2016), *A Comparison of Classifiers for Daily Activity Recognition Using Mobile Devices*, International Conference on Electrical, Electronic and Computing Engineering IcETRAN, Vol. 3, pp. MEI2.3.1-5, Zlatibor, Srbija, ISBN 978-86-7466-618-0.

[3] Simović, A. (2014), *Sistemi preporuke u e-trgovini*, Impact of the Internet on Business Activities in Serbia and Worldwide, Sinteza 2014, pp. 846-852, doi: 10.15308/SInteZa-2014-846-852, Beograd, Srbija, ISBN 978-86-7912-539-2.

[4] Staletić P., Simović A., Lutovac M. (2010), *Elektronska prodavnica korišćenjem open-source softvera*, Telekomunikacioni forum TELFOR, Vol. 18, pp. 1253-1256, Beograd, Srbija, ISBN 978-86-7466-392-9.

РАДОВИ У ЗБОРНИЦИМА НАЦИОНАЛНИХ СКУПОВА ШТАМПАНИ У ЦЕЛИНИ (М63):

[1] Simović A. (2017), *Big Data Analytics in Public Administration*, XLIV Symposium on Operational Research SYM-OP-IS 2017, Vol. 44, pp. 382-387, Zlatibor, Srbija, ISBN 978-86-7488-135-4.

- [2] Simović A., Čirović Z. (2016), *Sistem preporuke u elektronskoj trgovini u Big Data okruženju*, ETRAN 2016, Vol. 60, pp. RT4.5.1-4, Zlatibor, Srbija, ISBN 978-86-7466-618-0.
- [3] Simović A., Čirović Z. (2016), *Primena Hadoop okvira u analizi velikih skupova podataka*, XV međunarodni naučno-stručni Simpozijum INFOTEH, Vol. 15, pp. 593-598, Jahorina, ISBN 978-99955-763-9-4.
- [4] Čirović Z., Simović A. (2016), *Detekcija fizičkih aktivnosti pomoću akcelerometra, magnetometra i žiroskopa*, XV međunarodni naučno-stručni Simpozijum INFOTEH, Vol. 15, pp. 742-745, Jahorina, ISBN 978-99955-763-9-4.
- [5] Staletić N., Simović A. (2014), *Crowdsourcing u e-obrazovanju*, XLI Simpozijum o operacionim istraživanjima SYM-OP-IS 2014, Vol. 41, pp. 47-52, Divčibare, Srbija, ISBN 978-86-7395-325-0.
- [6] Staletić N., Simović A., Staletić P. (2014), *Comparative analysis of the Internet services application in the 2013th as compared to 2011*, International Scientific Conference Science and Higher Education in Function of Sustainable Development – SED 2014, Vol. 7, pp. 2.62-2.66, Uzice, Serbia, ISBN 978-86-83573-43-1.
- [7] Staletić N., Staletić P., Simović A. (2014), *Sigurnost Tor mreže u zaštiti identiteta na Internetu*, XIII međunarodni naučno-stručni Simpozijum INFOTEH, Vol. 13, pp. 913-917, Jahorina, ISBN 978-99955-763-3-2.
- [8] Simović A., Štrbac-Savić S., Staletić N., Čoko D. (2014), *Komparativna analiza Open Source CM sistema primenljivih u e-trgovini*, XIII međunarodni naučno-stručni Simpozijum INFOTEH, Vol. 13, pp. 743-748, Jahorina, ISBN 978-99955-763-3-2.
- [9] Staletić N., Staletić P., Simović A. (2013), *Funkcionalnosti yahoo groups servisa na studijskom programu elektronsko poslovanje*, XL Simpozijum o operacionim istraživanjima SYM-OP-IS 2013, Vol. 40, pp. 81-86, Zlatibor, Srbija, ISBN 978-86-7680-286-9.
- [10] Štrbac-Savić S., Simović A., Staletić N., Rankov S. (2013), *E-prodavnica knjiga Visoke škole elektrotehnike i računarstva u Beogradu*, XII međunarodni naučno-stručni Simpozijum INFOTEH, Vol. 12, pp. 896-899, Jahorina, ISBN 978-99955-763-1-8.
- [11] Staletić N., Martinović D., Simović A., Staletić P. (2013), *Mailing list – Support the Educational Process in Higher Education*, International May Conference on Strategic Management, IMKSM2013, pp. 920-927, Bor, Serbia, ISBN 978-86-6305-006-8.
- [12] Simovic, A., Lutovac, M. (2012), *Strategic Management Decisions on E-Commerce Solutions for Small Companies*, In May Conference on Strategic Management MKSM2012, Vol. 8, pp. 208-214, ISBN 978-86-80987-96-5.

11.4. ИЗЈАВА О АУТОРСТВУ

Име и презиме аутора: **АЛЕКСАНДАР М. СИМОВИЋ**

Број индекса: **5021/2016**

Изјављујем

да је докторска дисертација под насловом:

МОДЕЛ ПАМЕТНЕ БИБЛИОТЕКЕ ЗАСНОВАН НА *BIG DATA* ТЕХНОЛОГИЈАМА

- резултат сопственог истраживачког рада;
- да дисертација у целини ни у деловима није била предложена за стицање друге дипломе према студијским програмима других високошколских установа;
- да су резултати коректно наведени и
- да нисам кршио ауторска права и користио интелектуалну својину других лица.

Потпис аутора

У Београду, _____

11.5. ИЗЈАВА О ИСТОВЕТНОСТИ ШТАМПАНЕ И ЕЛЕКТРОНСКЕ ВЕРЗИЈЕ ДОКТОРСКОГ РАДА

Име и презиме аутора: **АЛЕКСАНДАР М. СИМОВИЋ**

Број индекса: **5021/2016**

Студијски програм: **ИНФОРМАЦИОНИ СИСТЕМИ И КВАНТИТАТИВНИ МЕНАЏМЕНТ, ИЗБОРНО ПОДРУЧЈЕ ЕЛЕКТРОНСКО ПОСЛОВАЊЕ**

Наслов рада:

МОДЕЛ ПАМЕТНЕ БИБЛИОТЕКЕ ЗАСНОВАН НА *BIG DATA* ТЕХНОЛОГИЈАМА

Ментор: **ПРОФ. ДР АЛЕКСАНДРА ЛАБУС**

Изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао ради похрањивања у **Дигиталном репозиторијуму Универзитета у Београду**.

Дозвољавам да се објаве моји лични подаци везани за добијање академског назива доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

Потпис аутора

У Београду, _____

11.6. ИЗЈАВА О КОРИШЋЕЊУ

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

МОДЕЛ ПАМЕТНЕ БИБЛИОТЕКЕ ЗАСНОВАН НА *BIG DATA* ТЕХНОЛОГИЈАМА

која је моје ауторско дело.

Дисертацију са свим прилозима предао сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигиталном репозиторијуму Универзитета у Београду и доступну у отвореном приступу могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (*Creative Commons*) за коју сам се одлучио.

1. Ауторство (CC BY)
2. Ауторство – некомерцијално (CC BY-NC)
3. Ауторство – некомерцијално – без прерада (CC BY-NC-ND)
4. Ауторство – некомерцијално – делити под истим условима (CC BY-NC-SA)
5. Ауторство – без прерада (CC BY-ND)
6. Ауторство – делити под истим условима (CC BY-SA)

Потпис аутора

У Београду, _____
