

УДК 004.6
004.6:02

doi: 10.19090/cit.2020.37.82-92
Оригинални научни рад

Технологија великих података (Big Data) и примена у будућој библиотечкој пракси

Јелена Цуца
jelena.angelovski@gmail.com

Сажетак

Са дигиталном револуцијом долази до масовне продукције свих врста података, структурираних и неструктурираних – и једно од решења по питању обраде било ког типа информација, текстуалног, аудио или видео, дефинише се развојем технологије великих података (Big Data). Њу карактерише велики обим и разноврсност података које је могуће обрадити у кратком временском периоду. Употребна вредност добијених резултата огледа се у истинитости и вредности информација које пружају увид у датасферу будућности. Описана технологија се примењује у свим областима развоја, од друштвених, природних и техничких наука, преко администрације, здравља, образовања, до привреде и економије. Коришћење Big Data у библиотекама представља нов начин да се корисницима пруже тражена знања и да се њихово пословање прилагоди дигиталном окружењу.

Тема великих података у стручним библиотечарским часописима у Србији је мало заступљена. Истраживањем стручне литературе на енглеском језику о вези ове технологије и библиотечарства, долазимо до закључка да ни у свету не постоји довољан број емпиријских студија. Библиотечарима недостаје основно разумевање практичне примене технологије великих података, иако су свесни њених предности. Циљ рада је да се она приближи стручној и научној јавности и укаже на могуће примене и изазове за библиотечку и информациону праксу.

Кључне речи: велики подаци, Big Data, информација, дигиталне информације, управљање дигиталним информацијама, анализа података, обрада података, постиндустријско друштво, дигитално друштво, библиотеке 4.0

Увод

Међународна корпорација за податке (*International Data Corporation – IDC*) предвиђа да ће до 2025. године глобални дигитални корпус превазићи 175 зетабајта, што је око 175 милијарди терабајта, док ће просечна особа повезана на интернет имати једну интеракцију са дигиталним уређајем на сваких 18 секунди.¹ Масовна продукција дигиталних информација повећава се из године у годину. Њихова организација представља велики изазов ако се у обзир узму и аналогни подаци који се дигитализују накнадно. Развијена свест о проблему неуређених информација, које као такве немају никакву употребну вредност, већ остају непримећене, или још горе – представљају терет за информационе системе и системе за складиштење, инспирисала је научну и стручну јавност да покуша да реши проблем. Правац деловања текао је ка једноставнијим начинима на које информације могу бити преточене у знање.

Уз напредак технологије који је донео брзе процесоре и велике меморије такозваних суперкомпјутера, базе података и њихово претраживање представљале су користан алат за управљање информацијама. Захваљујући њима омогућено је давање жељеног резултата претраге на основу података који су уношени у базу. Али они су могли бити сортирани, претраживани и анализирани само уколико су структурирани и једнаки по формату. Ако се посматра, на

¹ David Reinsel, John Gantz, and John Rydning, *Data Age 2025: The Evolution of Data to Life-Critical* (Framingham: International Data Corporation, 2017), преузето 25. 5. 2018, <https://www.seagate.com/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf>.

пример, каталожка база у некој библиотеци, записи у њој састоје се само од текстуалних информација. Уколико би неко желео да поред каталога претражује и колекције дигитализоване библиотечке грађе (фотографија, видео-записа или комплексних научних формула) алгоритми за претрагу би морали да се прилагоде, како би се претрага одвијала над различитим форматима докумената под једнаким критеријумима. Поред формата, изазов за информатичке и информационе стручњаке је представљала и величина ових неструктурираних података, који заузимају много више меморије од текстуалних. Такође, развојем друштвених мрежа, повећала се продукција садржаја. Како је наведено на блогу microfocus.com, у протеклој години су корисници друштвених медија дневно креирали 1.209.600 нових информација. Број особа које се служе интернетом порастао је од 2014. до 2019. године за 83%, а сваке године друштвени медији имају 300 милиона нових корисника.²

Алтруистичка и интердисциплинарна димензија библиотекарства кроз историју је подразумевала напредак струке у корак са развојем друштва. Свака новина која би могла да се искористи у сврху ефикаснијег и подробнијег ширења знања била је добродошла. Штампане каталожких листића, аутоматизована каталогизација, међубиблиотечка размена, обједињени каталози и други онлајн сервиси, као и дигитализација, само су неки од примера унапређења библиотечко-информационе струке на темељима технолошких иновација. Дигитализација библиотечког материјала представља важан сегмент развоја, јер уводи неструктуриране податке у област претраге. 3Д модели, видео-записи и други формати у којима су информације претраживе путем дигитализованог фонда, библиотекарима нуде могућност квалитетнијег одговора на захтев корисника, али представљају нов професионални изазов. Технологија великих података пружа могућност информационим и библиотечким стручњацима да ефикасно руководе колекцијама информација које су све присутније у корисничким захтевима.

Да би се успоставила примена ове технологије у библиотекарству, неопходно је да се библиотекарски едукују о појму и функцијама које пружа и начинима на који може унапредити њихов рад. Важно је добро проценити предности и мане њеног коришћења у сврху развоја струке.

Проучавајући литературу на тему великих података у библиотекарству открили смо само један научни рад из ове области у Србији.³ У њему се посредно говори о овој технологији као могућој иновацији у домену дигиталне, научне и академске писмености. Озбиљан недостатак прилога посвећених овој теми указује на потребу да се она детаљније обради и приближи стручној јавности.

Велики подаци – 3В и 5В концепти

Велики подаци дефинишу се на различите начине, у зависности од тога да ли се полази са техничких, технолошких, научних становишта, или се овај појам посматра кроз социолошки, аналитички или економски утицај који има на друштво. Појам је толико уопштен, да се не може тачно претпоставити ко га је први употребио, посебно у контексту у којем се употребљава данас. Модерно тумачење подразумева велики сет структурираних и неструктурираних дигиталних података који се обрађује великом брзином у циљу предвиђања будућих акција. Управо придев „велики“ јесте релативан појам у овом сегменту, па је стога немогуће формирати јединствену дефиницију. Међутим, неколико њих се може узети као приближно тачне јер свака из своје перспективе описује суштину Big Data.

² Jeff Schultz, "How much data is created on the internet each day", Micro Focus, 8. 6. 2019, preuzeto 14. 10 2019, <https://blog.microfocus.com/how-much-data-is-created-on-the-internet-each-day/>.

³ Tibor Koltay, "Big data, big literacies?", *Читалиште: научни часопис за теорију и праксу библиотекарства* Година XIII, број 24 (Мај 2014): 3–8, преузето 20. 5. 2020, https://citaliste.rs/casopis/br24/koltay_tibor.pdf.

Уопштена дефиниција која карактерише појам у довољној мери, али не прецизира о којим капацитетима је реч, наводи да су велики подаци они који премашују капацитет обраде конвенционалних система база података.⁴ Уколико се сагледа експоненцијални раст дигиталних информација, закључује се да ће обим или капацитет који се тренутно сматра великим у будућности бити стандард, а у прошлости је представљао само теоријску претпоставку. Такође, фраза „конвенционални системи база података“ представља променљиву категорију условљену технолошким напретком. Ова поједностављена дефиниција је у широкој употреби, посебно у домену техничке обраде информација, јер указује на разлику досадашњег организовања података структурисаних форми и новог вида организације неструктурисаних.

За библиотекарe би дефиниција Дага Лејнија из 2001. године била најрелевантнија као полазна тачка за проучавање ове технологије и њене примене у пракси. Аутор уводи концепт „3В“ (3V) за дефинисање великих података, који касније улази у широку употребу. Лејни користи „3В“ фразу за дефинисање три карактеристике великих података:

1. Обим (*Volume*)
2. Брзину (*Velocity*)
3. Разноврсност (*Variety*).

„3В“ дефинише Big Data као пакете података великог обима, брзине и разноврсности који захтевају економичне и иновативне процесе обраде информација и доприносе бољем увиду и доношењу одлука.⁵ У овом случају тај обим подразумева егзабајте, зетабајте и бронтобајте.⁶ Раст обима дигиталних информација једна је од битних карактеристика данашњег друштва и свакодневни живот у 21. веку ослања се, у великој мери, на њих. Послови се обављају преко интернета, а дигитални документи и потписи су постали званично валидни. Слободно време проводи се на друштвеним мрежама, што такође повећава продукцију дигиталних података. У првом кварталу 2018. године, просечни број активних корисника Твитера на месечном нивоу јесте 336 милиона.⁷ Примера ради, сабрана дела Вилијама Шекспира састоје се од 5,5 милиона карактера, а иста количина текста твитује се на сваких 11,6 секунди.⁸

Друга карактеристика Лејнијевог 3В концепта, брзина, односи се на више процеса: брзину којом се стварају нове информације, којом се оне шире виртуелним простором до других корисника и брзину којом се обрађују. Вирусни садржаји, видео-записи, фотографије, текстови и други формати се у реалном времену шире, преко интернета, веб-страницама и друштвеним мрежама и представљају типичан пример моћи брзине ширења података.

Треће „В“ из дефиниције представља разноврсност која се огледа кроз структуру података. Једноставно је разврстати текстуалне податке исте врсте, као што су алфанумерички низови, док је разноврсност у смислу видео, аудио и фото-формата за традиционалну обраду комплексније, ако не и немогуће. Овде се сви садржаји могу разврстати и анализирати без обзира на формат. Структурирани подаци могу се поредити са, на пример, телефонским имеником, а неструктурисани са кутијом у којој се чувају успомене као што су споменари, фотографије, видео и аудио-касете, писма и разгледнице.

Након развоја примене великих података, на Лејнијеву дефиницију додати су још и

4. Истинитост (*Veracity*)
5. Вредност (*Value*)

⁴ Edd Wilder-James, *What is big data: an introduction to the big data landscape*, O'Reilly Media, Inc., preuzeto 22. 5. 2018, <https://www.oreilly.com/ideas/what-is-big-data>.

⁵ Doug Laney, *3D Data Management: Controlling Data Volume, Velocity, and Variety* (Stamford: META Group Inc., 2001), preuzeto 25. 5. 2018, <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>.

⁶ 1EB (егзабајт) = 10¹⁸ бајтова, 1ZB (зетабајт) = 10²¹ бајтова, 1BB (бронтобајт) = 10²⁷ бајтова.

⁷ Statista, "Number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2018 (in millions)", preuzeto 21. 5. 2018, <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>.

⁸ AnnaLissa Sison, "How much data does Twitter store daily?", Quora Inc., preuzeto 21. 5. 2018, <https://www.quora.com/How-much-data-does-Twitter-store-daily>.

Дефиниција и концепт „5В“ мењали су се са развојем идеја о примени великих података. Научници, теоретичари и пословни људи додавали су још особина у такозвани „В“ концепт.¹⁰ Не може се са сигурношћу рећи ко је први додао истинитост и вредност, нити се може тврдити да је овај концепт дефинитиван (услед мењања технологије и области примене постоји могућност да се и „5В“ концепт измени или прошири), али један од првих заговорника додавања истинитости био је технолошки гигант ИБМ,¹¹ а вредност међу првима промовише Бернард Мар.¹² Потреба да се ове особине додају у дефиницију постоји како би се концепт Big Data прецизније одредио.

Истинитост указује на информације које се разликују у смислу квалитета, величине, релевантности и тачности. Коришћењем технологије великих података, увидело се да постоје и такозвани „запрљани подаци“, као што су погрешни метаподаци (тагови), намерно пласиране неистине, непроверене информације и други. Иако грешке које узрокује људски фактор могу да се предвиде програмирањем алгоритама за обраду великих података, људска намера не може, и постоји могућност добијања нетачних резултата. Вредност коју друштво добија захваљујући овој технологији јесте квалитативни резултат анализе података.

Питер Мел, компјутерски научник Националног института за стандарде и технологију (NIST – National Institute of Standards and Technology, U. S. Department of Commerce, Maryland, United States) сматра да се о појму велики подаци говори када запремина података, брзина набавке или представљање података ограничавају способност извршења ефикасне анализе користећи традиционалне релационе приступе, или захтевају коришћење значајног хоризонталног скалирања за ефикасну обраду.¹³

Социолошку дефиницију дају Ричардс и Кинг, ослањајући се на теорију Мајер-Шенбергера и Цукиера која дефинише велике податке као технолошке операције које се могу обавити у великом обиму, али не и у малом, да би се извукли увиди или створили нови облици вредности, на начине који мењају тржишта, организације и односе међу људима и влада – и у ширем, друштвеном смислу полемишу о великим подацима као појави која узрокује повећану моћ институционалне свести.¹⁴

Поред наведених, постоје још многе дефиниције појма велики подаци и чини се, што је термин једноставнији, то је теже објаснити његову специфичну примену. Оно што је заједничко за све дефиниције, јесу изазови везани за количину и величину података, као и за брзину и квалитет обраде.

Према тврдњама Међународне корпорације за податке (International Data Corporation), постоје и дигитални записи који су недоступни за обраду путем ове технологије. Процеси који претходе употреби информација подразумевају њихово обележавање, односно таговање и касније, анализу. У 2012. години 23% информација из датасфере (око 643 егзабајта) би било корисно уколико би оне биле таговане и анализиране. Прошле године, према истом извору, количина њихове искоришћености износила је 40%.¹⁵ Скривена знања у осталих 60% могу

⁹ Bernard Marr, “Big Data: The 5 Vs Everyone Must Know”, LinkedIn, March 6, 2014, preuzeto 12. 4. 2020, <https://www.linkedin.com/pulse/20140306073407-64875646-big-data-the-5-vs-everyone-must-know>.

¹⁰ Eileen McNulty, “Understanding big data: the seven V’s”, Dataconomy, May 22, 2014, preuzeto 12. 4. 2020, <https://dataconomy.com/2014/05/seven-vs-big-data/>; Neil Biehn, “The missing V’s in big data: viability and value”, Wired, preuzeto 12. 4. 2020, <https://www.wired.com/insights/2013/05/the-missing-vs-in-big-data-viability-and-value/>; Alex Popescu and Ana-Maria Bacalu, “Big Data: Volume, Velocity, Variability, Variety”, June 9, 2011, preuzeto 12. 4. 2020, <https://nosql.mypopescu.com/post/6361838342/bigdata-volume-velocity-variability-variety>.

¹¹ Biehn, “The missing V’s in big data”.

¹² Marr, “Big Data”.

¹³ Frank Konkel, “Sketching the big picture on big data”, FCW, 1105 Public Sector Media Group, preuzeto 21. 5. 2018, <https://fcw.com/articles/2013/04/15/big-experts-on-big-data.aspx?m=1>.

¹⁴ Neil M. Richards and Jonathan H. King, “Big Data Ethics”, *Wake Forest Law Review*, May 2014, 395, preuzeto 24. 5. 2018, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2384174.

¹⁵ Jovan Milenkovic, “30 Eye-Opening Big Data Statistics for 2020: Patterns Are Everywhere”, Kommando Tech, December 4, 2019, preuzeto 12. 4. 2020, <https://kommandotech.com/statistics/big-data-statistics/>.

допринети значајном развојном помаку друштва, уколико се ефикасно организују и искористе на прави начин. Велики рад и улагања у области Big Data допринеће њиховом бољем искоришћењу, а информације које се креирају и насељавају дигитални универзум већ повећавају количину корисних сазнања.¹⁶

Са социолошког становишта гледано, Big Data представља основу за грађење новог друштвеног поретка који је заснован на развијеној свести о свету у којем живимо и правцима у којима се он развија. Ако постоји свест о подацима које је могуће употребити у медицинске сврхе, заједница ће очекивати развој медицине. Ако постоји свест о онима који се узимају са друштвених мрежа, камера за видео-надзор и са персоналних рачунара, таблета и телефона, грађани ће имати свест о могућности формирања безбеднијег друштва. То би омогућило и ефикасније полицијско деловање које ће се заснивати на предвиђањима криминалних радњи и њиховим спречавањима пре него што се десе. Ако су државни подаци транспарентни, заједница ће имати јачи осећај за демократију. На крају, ако се технологија великих података користи на исправан начин и ако има за циљ развој и напредак, друштво ће имати поверења у технологију предвиђања и решавања проблема на овај начин. Пример добре праксе јесте Естонија у којој је 99% државних сервиса онлајн, исти је постотак грађана који имају електронску личну карту, а 43,8% становника је 2019. гласало на парламентарним изборима путем интернета, за разлику од 2007, када је само 5,5% грађана користило ову опцију.¹⁷

Према речима професора Андреаса Вајгенда са Стенфорд универзитета, информације су нова нафта; као и она, морају бити рафинисане пре коришћења.¹⁸ Овом констатацијом могло би се заокружити поимање Big Data, али и тежине и важности коју информације данас имају.

Употреба великих података

Новине које долазе овим начином анализе информација тичу се извора података, типова, обима и квалитета, као и критеријума за сакупљање, обраду и класификацију које је неопходно одредити унапред и на крају тумачења и коришћења добијених резултата.

Процес анализе података следи њихово прикупљање и избор и подразумева примену алгоритама и различитих компјутерских алата приликом селектовања и обраде. За њих је карактеристично да имају способност прилагођавања својих параметара у односу на информације које примају и на тај начин стварају нове алгоритме који се даље баве деловима података од интереса. Овако напредни модели представљају суштину револуционарне особине технологије Big Data јер превазилазе нелинеарност информација чија хетерогеност захтева много комплекснију поставку приступа подацима ради жељеног резултата. Коришћење технологије великих података смањује људске производе у овом процесу и тиме значајно скраћује временски оквир у производњи, сакупљању и преносу информација, што даље омогућава више времена и људских ресурса за анализу и тумачење података, као и за доношење одлука базираних на добијеним садржајима.

Све чешћа примена технологије великих података искристалисала је потребу индустрије за одређеним профилом стручњака који може бити кључан када је анализа великих података укључена у пословни процес. На глобалном нивоу, дефицитарно занимање јесте информациони стручњак (information specialist). То је особа, или тим људи, која анализира резултате и најважније – даје предлог како даље деловати. Наука о информацијама постала

¹⁶ John Gantz and David Reinsel, *The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East* (Framingham: International Data Corporation, 2012), преузето 25. 5. 2018, <https://www.emc.com/collateral/analyst-reports/idc-digital-universe-united-states.pdf>.

¹⁷ E-Estonia, "We have built a digital society and we can show you how", E-Estonia, преузето 14. 4. 2020, <https://e-estonia.com/>.

¹⁸ UN Global Pulse, *Big Data for Development: Challenges and Opportunities* (New York: UN Global Pulse, 2012), преузето 20. 5. 2018, <http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf>.

је веома популарна и широм света развијају се програми за академско образовање специјалиста у овој области. Пресудно за информационог стручњака јесте сагледавање потреба из перспективе понуђених информација, аналитичко расуђивање, креативност и, пре свега, константно образовање о новим применљивим технологијама за обраду и анализу података.

Из поља науке требало би истаћи пример највећег и најснажнијег акцелератора честица на свету. ЦЕРН-ов гигант (CERN – Conseil Européen pour la Recherche Nucléaire), велики хадронски сударац је 29. јуна 2017. године прошао прекретницу од 200 петабајта података који су стално архивирани у њиховим библиотекама. Одакле они долазе? Честице се сударају у акцелератору отприлике милијарду пута у секунди, стварајући око један петабајт података судара у секунди. Међутим, постојећи рачунарски системи не евидентирају све и задржавају само најзначајније информације. Филтрирани подаци се затим сабирају у ЦЕРН-овом центру, где се обавља њихова иницијална реконструкција и где се копирају за дуготрајно архивирање. Чак и након драстичног смањења података изведених у експериментима, ЦЕРН обрађује у просеку један петабајт података дневно.¹⁹

Примена великих података у библиотекарству

Питање увођења нових технологија и трансформација библиотечких услуга наводе на размишљања о будућој улози самих библиотека и библиотекара. Област библиотекарства и информационих наука има традицију организовања, управљања, преузимања, прикупљања, описивања и очувања информација. Иако су оне у прошлости биле у облику кодираних врста научних комуникација, попут монографија и чланака из часописа, сада су прешле на инклузивније студије свих врста информација, без обзира на облик. Ово проширује област од оне која се углавном фокусира на текстуалне научне публикације, до поља која проучавају податке које стварају истраживачи и јавност.²⁰

Као и свака нова и револуционарна технологија, тако и ова великих података има својих мана. Велики подаци се називају тако не само због „5В“ карактеристика, већ и због тежине и богатства метаподацима. Један твит није само низ од највише 280 карактера, он представља много више од тога. Од контекста зашто, када и одакле је написан, од информација ко га је написао и даље, укрштањем података добијених на тај начин са банковним подацима или историјом интернет претрага, рачуна из ресторана те особе, може се сакупити велики број информација. Оне се могу анализирати и произвести нове, другим речима, могу пружити ново знање. Са друге стране, могу се и злоупотребити. Коришћење сазнања у погрешне сврхе није новост, али са растом обима обрађених података, расте и шанса да се лични подаци користе без сагласности особе која их је дала. Ситуација са пандемијом КОВИД-19 показала је како држава може да повреди права грађана. Овакви примери оправдано буде неповерење и страх грађана у дигитално обрађено знање.²¹

Продукција и могућност једноставног и брзог ширења информација захваљујући интернету, преносивим уређајима, свакодневним иновацијама из области сакупљања и обраде података,

¹⁹ Méli­ssa Gaillard and Stefania Pandolfi, “CERN Data Centre passes the 200-petabytes milestone”, CERN Data Centre, July 6, 2017, preuzeto 25. 5. 2018, <http://cds.cern.ch/record/2276551>.

²⁰ Cassidy R. Sugimoto, Ying Ding, Mike Thelwall, “Library and Information Science in the Big Data Era: Funding, Projects and Future (a panel proposal)”, American Society for Information Science and Technology, January 24, 2013, preuzeto 20. 5. 2018, <https://asistdl.onlinelibrary.wiley.com/doi/epdf/10.1002/meet.14504901187>.

²¹ Stuart A. Thompson and Charlie Warzel, “Twelve Million Phones, One Dataset, Zero Privacy”, *The New York Times*, December 19, 2019, preuzeto 25. 6. 2020, <https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html>; Cheng-hu Zhou et al., “COVID-19: Challenges to GIS with Big Data,” *Geography and Sustainability* Volume 1, issue 1 (March 2020): 77–87, preuzeto 25. 6. 2020, <https://doi.org/10.1016/j.geosus.2020.03.005>; Yves-Alexandre de Montjoye et al., “Unique in the Crowd: The privacy bounds of human mobility”, *Scientific Reports* Volume 3, Article number: 1376 (March 25, 2013): 1–5, preuzeto 25. 6. 2020, DOI: 10.1038/srep01376.

представљају велику шансу за библиотеке да оствари своје циљеве. Технологија Big Data захтева знање из области управљања информацијама и то је тачка где се потенцијал библиотека и ове технологије сусрећу. Знање и стручност библиотекарa из области великих података омогућава доступност најновијим информацијама и подацима у различитим форматима брже него у случају традиционалних извора знања. Уз познавање правне регулативе интелектуалних и ауторских права, они су у могућности да корисницима правовремено и на одговоран начин дају грађу на коришћење.

У светлу нових трендова, оно што разликује све популарније информационе стручњаке од традиционалних библиотекара јесу активности везане за примену нових технологија и управљање савременим изворима знања која постају дигитално доступна. Тако ИТ стручњак ради на управљању корисничким лиценцама и лиценцираним подацима које купује од издавача, пружа саветодавне услуге у вези са коришћењем различитих извора информација, пружа савете из области заштите ауторских и интелектуалних права, као и из области заштите приватности података и уопштено, представља извор знања о управљању структурираним и неструктурираним подацима. Са друге стране, Према *Онлајн речнику библиотекарства и информационог наука (ODLIS – Online Dictionary for Library and Information Science)* Џоане М. Риц, библиотекар се дефинише као професионално обучена особа одговорна за бригу о библиотеци и њеним садржајима, укључујући избор, обраду и организацију материјала и пружање информација, упутстава и услуга позајмљивања, како би се задовољиле потребе корисника. У онлајн окружењу улога библиотекара јесте да управља и посредује у приступу информацијама које могу постојати само у електронском облику.²² Суштински, нема разлике у врсти посла традиционалног библиотекара и информационог стручњака. То су два назива која су се раздвојила само услед динамичне промене извора информација којима информациони кадар управља.

У циљу пружања што квалитетније услуге својим корисницима, односно грађанима, јавне библиотеке пролазе кроз константну реформу. Убрзани напредак дигиталног друштва захтева још напредније пословање. Доступност података, упознавање са новим технологијама, заштита приватности, класификација и чување дигиталног фондуса представљају изазов за сваког појединца, за друштво у целини и за библиотеке. Посебан изазов стоји пред јавним библиотекама због обима фонда и великог броја корисника, као и разноврсног спектра захтева на које морају одговорити. Јавне библиотеке могу пружити адекватну обуку шире јавности о значају концепта Big Data. Не сналазе се сви корисници јавних библиотека у дигиталном окружењу, тако да би едукација којој би свако заинтересован имао приступ била од користи да се објасне предности нових технологија. Нова сазнања и објашњења би умањила неповерење грађана у дигиталну обраду података. Осим коришћења информација у смеру од извора до корисника, Big Data омогућавају библиотекари да, уз обраду података о корисницима, пруже квалитетнију услугу и бољи сервис. У велике податке спадају, како је раније у тексту наведено, и они о понашању и навикама појединаца. Библиотеке било ког типа, од јавних, до специјалних, могу их проучити и на тај начин креирати план за будуће пословање. На тај начин оне не само да прикупљају, класификују, чувају и корисницима пружају знање на коришћење, већ и активно учествују у креирању нових сетова информација и на тај начин у потпуности користе технологију великих података. Библиотеке које су на овај начин у могућности да предвиде потребе својих корисника и заједнице, могу брже да доносе исправне одлуке, политику, планове и програме рада.

Потребе корисника се мењају и у академским библиотекама. Чланци из стручних и научних часописа нису више најтраженији извори информација. Данас су најрелевантнији подаци

²² Joan M. Reitz, *ODLIS – Online Dictionary for Library and Information Science*, ABC Clío, preuzeto 14. 4. 2020, https://products.abc-clio.com/ODLIS/odlis_about.aspx.

они који долазе из најновијих истраживања, са тек одржаних научних конференција, комплексне формуле, компјутерски програми и слично који имају заједничку особину – да не спадају у стандардне, лако претраживе структуриране податке.

Специјалне библиотеке и информациони центри све више буџете усмеравају на куповину база података, извештаја који садрже неструктуриране информације, или то јесу по својој форми, и других извора информација. Научни и стручни часописи најчешће не садрже актуелне податке до времена када стигну у библиотеку и прођу каталожку обраду. Стога су кратке форме, на бази плаћања лиценци за коришћење, које брзо стижу до корисника или им се уз помоћ налога може у сваком моменту приступити, постале предмет интересовања ових библиотека. Њих карактерише мали број корисника заинтересованих за уже научне дисциплине. Библиотекари морају имати знања из те области, уз своју експертизу из информационих наука. Специјалне библиотеке фокусирају се на референсне збирке, најновије објављене информације и њихову ефикасну дисеминацију корисницима. То подразумева управљање базама података, проналажење најефикаснијег начина за чување дигиталног фонда, али и поштовање ауторских права и лиценци. Осим по захтеву, подаци се шаљу корисницима и на иницијативу библиотекара. Задатак запосленог у оваквој библиотеци јесте да креира профиле корисника према ужој области интересовања и да доставља правовремене и проверене изворе информација. Технологија Big Data омогућава сакупљање свих типова података, њихов избор и проверу. Ефикасност електронске обраде је у поређењу са мануелном неупоредива, али кључан фактор представља знање библиотекара о овој технологији, односно, квалитет знања које се добија као резултат обраде зависи од инпута које задаје библиотекар.

На основу поменутих могућности коришћења података комерцијалног типа, где издавачи наплаћују лиценце за свој садржај и података у вези са корисницима библиотека, природно се јавља брига о правној регулативи и заштити појединаца и ауторских права издавача. Управо библиотеке у овом поретку представљају комуникациони канал путем којег знање, у облику тражених података, долази до корисника, уз пуно поштовање правних и етичких стандарда.

Овадиа примећује да једна од улога библиотекара може бити да посредују у размени сецова великих података, да штите права издавача, али и да раде на доступности информација. Аутор даље наглашава да је за библиотекарске веома важна заштита права корисника и да из тог разлога информациони стручњаци могу деловати и као демонстратори и заговорници техника за заштиту њихових права.²³

Нох претпоставља да ће у будућности библиотеке 4.0²⁴ бити системи који неће само пружати тражене информације корисницима, него ће и саме анализирати податке и креирати нове. Наводи да ће оне бити окружења која обједињују платформе, сервисе и велике изворе информација са интернета који ће омогућавати библиотекарима, корисницима и машинама да сарађују. Овај симбиотички веб представља технологију која дозвољава машинама и људима да пишу, читају, извршавају и закључују истовремено. Интелигентна библиотека 4.0 биће субјект који мисли, доноси одлуке и пружа услуге уз помоћ сопствених процена.²⁵

Изазова које пред библиотекарство ставља примена технологије великих података има много. У зависности од подршке оснивача и управних тела, новине се у институције уводе брже или спорије. Један од изазова свакако може бити буџет библиотеке, односно расположива средства за улагање у нове инвестиције. Са кадровске стране, изазов може бити увођење

²³ Steven Ovadia, "The Role of Big Data in the Social Sciences," *Behavioral & Social Sciences Librarian* 32, 2 (Online Jun 10, 2013): 133, preuzeto 25. 5. 2018, <https://doi.org/10.1080/01639269.2013.787274>.

²⁴ Библиотека 2.0 представља библиотеку са активним учешћем корисника чије су активности засноване на коришћењу веба 2.0, а библиотека 3.0 уз коришћење семантичког веб-окружења омогућава машинску обраду података и употребу друштвених мрежа са циљем размене знања и међусобне сарадње.

²⁵ Younghee Noh, "Imagining Library 4.0," *The Journal of Academic Librarianship* 41 (October 1, 2015): 786–797, preuzeto 25. 6. 2020, <https://doi.org/10.1016/j.acalib.2015.08.020>.

промена пословања и додатне обавезе запослених, као и организација обуке за руковођење великим подацима. Додатно, библиотекари би морали да буду у току са технолошким иновацијама, трендовима у истраживању и обради података, као и са правним регулативама. Као и свака новина, и ова може наићи на наведене препреке. Оне се могу превазићи образовањем управљачког тела и запослених, са циљем представљања користи коју установа може имати. Тек када стручно особље у потпуности савлада управљање великим подацима, биће у могућности да га заговара и користи.

Тема великих података у стручним библиотечарским часописима у Србији није довољно заступљена. Истраживањем стручне литературе на енглеском језику о вези ове технологије и библиотечарства, долази се до закључка да ни у свету не постоји довољно емпиријских студија. Велики број радова бави се теоријским претпоставкама и предлозима за унапређење рада библиотека на овај начин. Кроз истраживање које је 2017. године спроведено у финским јавним библиотекама, закључено је да библиотечарима недостаје основно разумевање практичне примене технологије великих података, иако су свесни њених предности.²⁶

Закључак

Могућност примене технологије великих података простира се од науке и образовања, преко развоја економије, администрације, људског развоја, све до културе и уметности. Масовност и важност података које је донело дигитално доба инспирисало је научни свет да осмисли како да се та врста знања најефикасније искористи.

Уколико је први корак информационе револуције био дефинисан микропроцесорима и моћи компјутеризације, а други умрежавањем и моћи повезивања, трећи ће бити дефинисан подацима и моћи претпостављања.²⁷ Развој ће бити заснован на технологији која омогућава квалитетан и тачан унос, обраду и анализу великог броја података и давања резултата у реалном времену, односно коришћење сазнајног капацитета које данас информације поседују. При томе, требало би бити крајње опрезан у коришћењу приватних података, како би се заштитила права човека. Злоупотреба информација не може допринети бољем развоју друштва, уколико заједница изгуби поверење у добробит технологије зарад напретка. Њих би требало сакупљати, али никако по сваку цену.

Иако је данас још увек тешко применити ову технологију у библиотечкој пракси, због нивоа квалитета који се мора постићи и због недостатка информационих стручњака који је могу користити на прави начин, улагање у њен развој ће се исплатити јер ће бити могуће данашње знање проширити знањем које носе велики подаци. Иако се за библиотекара данас користи назив информациони стручњак и, иако се њихов значај и активности мењају уз употребу нових технологија, они и даље имају исти циљ – учешће у ефикасном чувању, организацији и коришћењу знања. Заједно са улагањем у развој технологије и развој кадра у библиотекама, мора се улагати и у развој правних оквира и етичког коришћења информација. Једино овако свеобухватним деловањем ће велики подаци бити у служби свих грађана.

²⁶ Ming Zhan and Gunilla Widén, "Public libraries: roles in Big Data," *The Electronic Library* Vol. 36, No. 1 (2018): 133–145, preuzeto 25. 6. 2020, DOI 10.1108/EL-06-2016-0134.

²⁷ Richards and King, "Big Data Ethics": 395.

Литература и извори:

1. Biehn, Neil. "The missing V's in big data: viability and value". Wired. Preuzeto 12. 4. 2020. <https://www.wired.com/insights/2013/05/the-missing-vs-in-big-data-viability-and-value/>.
2. De Montjoye, Yves-Alexandre, César A. Hidalgo, Michel Verleysen and Vincent D. Blondel. "Unique in the Crowd: The privacy bounds of human mobility". *Scientific Reports* Volume 3, Article number: 1376 (March 25, 2013): 1–5. Preuzeto 25. 6. 2020. DOI: 10.1038/srep01376.
3. E-Estonia. "We have built a digital society and we can show you how". E-Estonia. Preuzeto 14. 4. 2020. <https://e-estonia.com/>.
4. Gaillard, Mélissa and Stefania Pandolfi. "CERN Data Centre passes the 200-petabytes milestone". CERN Data Centre. Preuzeto 25. 5. 2018. <http://cds.cern.ch/record/2276551>.
5. Gantz, John and David Reinsel. *The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East*. Framingham: International Data Corporation, 2012. Preuzeto 25. 5. 2018. <https://www.emc.com/collateral/analyst-reports/idc-digital-universe-united-states.pdf>.
6. Koltay, Tibor. "Big data, big literacies?" *Čitalište: naučni časopis za teoriju i praksu bibliotekarstva* Godina XIII, broj 24 (Maj 2014). Preuzeto 20. 5. 2020. https://citaliste.rs/casopis/br24/koltay_tibor.pdf.
7. Konkel, Frank. "Sketching the big picture on big data". FCW, 1105 Public Sector Media Group. Preuzeto 21. 5. 2018. <https://fcw.com/articles/2013/04/15/big-experts-on-big-data.aspx?m=1>.
8. Laney, Doug. *3D Data Management: Controlling Data Volume, Velocity, and Variety*. Stamford: META Group Inc., 2001. Preuzeto 25. 5. 2018. <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>.
9. Marr, Bernard. "Big Data: The 5 Vs Everyone Must Know". LinkedIn. March 6, 2014, preuzeto 12. 4. 2020. <https://www.linkedin.com/pulse/20140306073407-64875646-big-data-the-5-vs-everyone-must-know>.
10. McNulty, Eileen. "Understanding big data: the seven V's". Dataconomy. May 22, 2014, preuzeto 12. 4. 2020. <https://dataconomy.com/2014/05/seven-vs-big-data/>.
11. Milenkovic, Jovan. "30 Eye-Opening Big Data Statistics for 2020: Patterns Are Everywhere". Kommando Tech. December 4, 2019, preuzeto 12. 4. 2020. <https://kommandotech.com/statistics/big-data-statistics/>.
12. Noh, Younghee. "Imagining Library 4.0." *The Journal of Academic Librarianship* 41 (October 1, 2015). Preuzeto 25. 6. 2020. <https://doi.org/10.1016/j.acalib.2015.08.020>.
13. Ovadia, Steven. "The Role of Big Data in the Social Sciences." *Behavioral & Social Sciences Librarian*, 32, 2 (2013): 130–134. Preuzeto 15. 4. 2020. <https://doi.org/10.1080/01639269.2013.787274>.
14. Popescu, Alex and Ana-Maria Bacalu. "Big Data: Volume, Velocity, Variability, Variety". June 9, 2011, preuzeto 12. 4. 2020. <https://nosql.mypopescu.com/post/6361838342/bigdata-volume-velocity-variability-variety>.
15. Reinsel, David, John Gantz and John Rydning. *Data Age 2025: The Evolution of Data to Life-Critical*. Framingham: International Data Corporation, 2017. Preuzeto 25. 5. 2018. <https://www.seagate.com/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf>.
16. Reitz, Joan M. *ODLIS – Online Dictionary for Library and Information Science*. ABC Clío. Preuzeto 14. 4. 2020. https://products.abc-clio.com/ODLIS/odlis_about.aspx.
17. Richards, Neil M. and Jonathan H. King. "Big Data Ethics". *Wake Forest Law Review* 3 (2014). Preuzeto 24. 5. 2018. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2384174.
18. Schultz, Jeff. "How much data is created on the internet each day". Micro Focus, 8. 6. 2019, preuzeto 14. 10 2019. <https://blog.microfocus.com/how-much-data-is-created-on-the-internet-each-day/>.
19. Sison, AnnaLissa. "How much data does Twitter store daily?". Quora Inc. Preuzeto 21. 5. 2018. <https://www.quora.com/How-much-data-does-Twitter-store-daily>.
20. Statista. "Number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2018 (in millions)". Preuzeto 21. 5. 2018. <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>.

21. Sugimoto, Cassidy R., Ying Ding and Mike Thelwall. "Library and Information Science in the Big Data Era: Funding, Projects and Future (a panel proposal)". American Society for Information Science and Technology. January 24, 2013, preuzeto 20. 5. 2018. <https://asistdl.onlinelibrary.wiley.com/doi/epdf/10.1002/meet.14504901187>.
22. Thompson, Stuart A. and Charlie Warzel. "Twelve Million Phones, One Dataset, Zero Privacy". *The New York Times*, December 19, 2019, preuzeto 25. 6. 2020. <https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html>.
23. UN Global Pulse. *Big Data for Development: Challenges and Opportunities*. New York: UN Global Pulse, 2012. Preuzeto 20. 5. 2018. <http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf>.
24. Wilder-James, Edd. *What is big data: an introduction to the big data landscape*. O'Reilly Media, Inc. Preuzeto 22. 5. 2018. <https://www.oreilly.com/ideas/what-is-big-data>.
25. Zhan, Ming and Gunilla Widén. "Public libraries: roles in Big Data." *The Electronic Library* Vol. 36, No. 1 (2018): 133–145. Preuzeto 25. 6. 2020. DOI 10.1108/EL-06-2016-0134.
26. Zhou, Chenghu, Fenzhen Su, Tao Pei, An Zhang, Yunyan Du, Bin Luo, Zhidong Cao, Juanle Wang, Wen Yuan, Yunqiang Zhu, Ci Song, Jie Chen, Jun Xu, Fujia Li, Ting Ma, Lili Jiang, Fengqin Yan, Jiawei Yi, Yunfeng Hu, Yilan Liao and Han Xiao. "COVID-19: Challenges to GIS with Big Data." *Geography and Sustainability* Volume 1, issue 1 (March 2020): 77–87. Preuzeto 25. 6. 2020. <https://doi.org/10.1016/j.geosus.2020.03.005>.

Big Data Technology and Its Application in Future Library Practice

Summary

The digital revolution brings the mass production of all types of data, both structured and unstructured. One of the solutions in terms of processing any type of information – text, audio, or video – lies in the development of big data technology. It is characterized by a large volume and variety of data that can be processed in a short period of time. The use value of the obtained results is reflected in the truthfulness and value of information that provides insight into the datasphere of the future. This technology is applicable in all spheres of the development, from the social, natural, and technical sciences, through administration, health, education, to the economy.

There are many challenges that the application of big data technology poses to librarianship. Depending on the support of the founders and governing bodies, innovations are introduced into institutions faster or slower. One of the challenges can certainly be the library budget, i.e. the available funds for new investments. From the personnel point of view, the challenge may be the introduction of business changes and additional obligations of employees, as well as the organization of training in big data management. Besides, librarians should keep abreast of technological innovations, trends in research and data processing, as well as legal regulations. Like any innovation, this one can also encounter certain obstacles. They can be overcome by educating the governing body and employees, in order to point out the benefits of this concept. Only when professional staff have fully mastered big data management will they be able to advocate and use it.

The aim of this paper is to bring big data technology closer to the professional and scientific public in Serbia and point out its possible applications, advantages, and challenges for library and information practices.

Keywords: big data, information, digital information, digital information management, data analysis, data processing, post-industrial society, digital society, libraries 4.0



Технологија великих података (Big Data)

и примена у будућој библиотечкој пракси by Јелена Цуца is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.