A Big Data smart library recommender system for an educational institution

Aleksandar Simović
Department of E-Business, Faculty of Organizational Sciences, University of Belgrade, Belgrade, Serbia

Received 30 June 2017
Revised 30 October 2017
17 January 2018
14 March 2018
Accepted 14 March 2018

Abstract

Purpose – With the exponential growth of the amount of data, the most sophisticated systems of traditional libraries are not able to fulfill the demands of modern business and user needs. The purpose of this paper is to present the possibility of creating a Big Data smart library as an integral and enhanced part of the educational system that will improve user service and increase motivation in the continuous learning process through content-aware recommendations.

Design/methodology/approach – This paper presents an approach to the design of a Big Data system for collecting, analyzing, processing and visualizing data from different sources to a smart library specifically suitable for application in educational institutions.

Findings – As an integrated recommender system of the educational institution, the practical application of Big Data smart library meets the user needs and assists in finding personalized content from several sources, resulting in economic benefits for the institution and user long-term satisfaction.

Social implications – The need for continuous education alters business processes in libraries with requirements to adopt new technologies, business demands, and interactions with users. To be able to engage in a new era of business in the Big Data environment, librarians need to modernize their infrastructure for data collection, data analysis, and data visualization.

Originality/value – A unique value of this paper is its perspective of the implementation of a Big Data solution for smart libraries as a part of a continuous learning process, with the aim to improve the results of library operations by integrating traditional systems with Big Data technology. The paper presents a Big Data smart library system that has the potential to create new values and data-driven decisions by incorporating multiple sources of differential data.

Keywords Libraries, Data analysis, Big Data, Data storage, Education, Recommender system

Paper type Technical paper

Introduction

Traditionally configured systems for data storage and analysis prevent libraries from achieving competitive advantages. Over the past decade, many academic libraries have struggled to shift the value and utility of collected data (Buckland, 2017). Library users have switched to online scholar sources of information, and academic libraries have lost their monopoly over the provision of scientific information (Chambers, 2013). Library users retrieve new data necessary for learning, evaluate new theories, or discover a new addition to knowledge. Each of these functions involves determining the specific knowledge, professional literature and other learning materials that may not be available in the library (Feisel and Rosa, 2005). Identifying and analyzing data beyond the library, through campuses and external aggregations, can develop effective services and systems bringing value to the wider institution (Showers, 2015). Large amounts of available data and the implications of differential resources increase the complexity of their collection even further (Showers, 2014). Library catalogs need to carry enough information about items and users preferences to have the capacity to determine a potentially ideal result and to respond adequately to the given query (Horstmann and Brase, 2016).

Nowadays, the cumulative increase in the volume of data from various sources relational and non-relational (stored in local and in the cloud environment) has given rise to the problem of providing an efficient library service to the users. According to

Library Hi Tech
Vol. 36 No. 3, 2018
pp. 498-523
© Emerald Publishing Limited
0737-8831
DOI 10.1108/LHT-06-2017-0131

Rubin (2017), educational institutions are in a period of transition in how they deliver their library services. New methods of content delivery allow e-libraries to substitute traditional face-to-face librarian recommendations. Recommendation methods need to adapt/respond to continually improving new technologies offering the possibility of new forms of delivery just like the nature and structure of educational library catalogs might well change so as to offer content of greater interest in line with the emergence of new economic models. These challenging conditions make a smart library an inescapable part of a modern library business system, which itself is an integral part of learning processes and the educational institution.

The approach used in this work is based on the idea that the integration of Big Data technologies into smart library data management ecosystem solves issues such as: how large amounts of data from different sources can be collected and connected, integrated and stored, and analyzed and visualized; and how to display the content of more interest to users through a recommender system?

In order to improve the process of meeting user needs in the continuous educational cycle, the proposed Big Data recommender system enables data integration from various sources (e.g. Learning Management Systems (LMS), University online bookstore, Internet of Things (IoT), data from social media networks, and traditional library) into the smart library, making the approach particularly suitable for application in educational institutions.

The main contributions of this paper are:

(1) Recommendation systems were made based on practical requirements as personalized e-services that have application in different domains. Existing recommender systems mainly focus on well-known approaches and they are reviewed in the next section. This paper proposes an integrated recommendation system that complements existing systems and provides a useful guide for librarians, practitioners, and researchers in developing a Big Data smart library model, and creating a new service that will improve the educational process.

(2) This paper provides a framework for an efficient application of four independent data sources into the Big Data ecosystem making the smart library an integrated part of the educational continuum.

(3) For each data set, it effectively identifies the specific details and requirements for an integrated recommendation with the aim of improving the results of library operations by merging the traditional library and information systems (ISs) of an educational institution with a Big Data framework. This will motivate and support researchers and practitioners to promote the popularization of this approach.

(4) It particularly suggests possible further research of integrated recommendation systems in the Big Data era with the proposition of developing a smart library suitable for an educational institution.

The motivation behind this study is the lack of a comprehensive survey in the field of Big Data smart library recommenders that approaches the issue from the perspective of an educational institution as a foundation of knowledge creation and dissemination in society.

The paper is organized as follows. First, it provides a review of the literature related to Big Data technologies and recommender systems with a comparison of traditional systems and other research fields in diverse domains where have found application. The next section presents the Big Data smart library model with the ultimate aim being the development of an integrated recommender system suitable for an educational institution with a detailed outline of its implementation. That is followed by an evaluation of the system and its results. The next section is a discussion. Finally, in the conclusion there is discussion which includes the limitations, theoretical and practical implications of this research and outlines future work.

Literature review

Nowadays we are confronted with a lot of different types of data which is generated at a high rate. According to Gartner, Big Data is classified as one of the most important strategic technology trends in 2017 (Panetta, 2016). International Data Corporation (IDC) in its study estimated that the overall growth of the digital universe by 2020 would reach 40ZB of data (Edjlali *et al.*, 2012). Research and scientific institutions, as well as companies in the public and private sector, generates large amounts of data on various topics (Fey *et al.*, 2008): climate change, video surveillance of transport infrastructure, history of patients in health institutions, purchase history from e-commerce stores, social behavior through interaction on social networks, etc. (Boyd and Crawford, 2012). This is the result of more and more devices being integrated into the business processes of organizations (Cukier and Mayer-Schoenberger, 2013).

Laney (2001) has defined the challenges that bring large amounts of data through the "3Vs". Volume – the total data set size and cumulative volume beyond the capacity of existing relational database management systems (RDBMS) to process it; velocity – the speed at which new data is created which represents dynamic data-use through the interaction of participants; and variety – different formats of incompatible and inconsistent data structures, unorganized and large, carrying information of importance waiting to be analyzed and used (Buyya *et al.*, 2016). In 2012, Gartner expanded the Big Data definition and declared that ISs need new forms of processing to allow improved decision making, gaining insights and optimizing business processes (Beyer and Laney, 2012). IBM added another V attribute to Laney's definition – veracity. This is known as the "4Vs" (Erl *et al.*, 2016). Zikopoulos *et al.* (2012) explained the reason behind IBM's additional V and introduced the dimension of accuracy in relation to the quality of data sources in facing Big Data initiatives. According to some researchers, there are as many as "10Vs" in Big Data analytics in scientific papers (Markus, 2015). Based on the historical growth rate of the overall data generated and data flow, Cisco claims that humanity entered the ZetaByte era in 2015 (Cisco, 2017).

Big Data technologies have great application in recommender systems. They have taken the place of widely applied tools in various domains of digital businesses (Philip Chen and Zhang, 2014). However, they also have great potential to be applied in library ISs and to contribute a better understanding of user needs by proposing content of interest. The goal of the recommendation system is to reduce information overload by extracting the most relevant content and information from a vast amount of data. Adomavicius and Tuzhilin (2005) presented an overview of content-based (CB), collaborative filtering (CF) and hybrid-based recommenders, describing the limitations of these approaches and discussed possible improvements. These approaches prognosticate the level of user interest or the usefulness of a particular item and rank them according to predicted values (Bernardes *et al.*, 2014). CB is based on the assumption that the usefulness of the item will be similar to those that the user has preferred in the past, while CF predicts the usefulness of a particular item for a user based on the evaluation of that item by other system users. The combination of these algorithms creates a hybrid-personalized system of recommendations which calculates both, the user's rating, and the function of the content (Herlocker *et al.*, 2000; Adomavicius and Tuzhilin, 2005; Lu *et al.*, 2015). Recommender systems have a common feature that is recognized as an important source of information to offer to users to rate items and post reviews according to their opinion (Lin, 2014). Lee *et al.* (2015) advocated the need for new techniques to reduce bias in movie ratings, raising questions about the reliability of ratings as an impartial quality indicator. They also found that a prior rating by an online community as crowd vs friends can have a varying impact on subsequent user's ratings. Gao *et al.* (2018) investigated the influence of cultural factors on the users' online rating behavior focusing on how cultural values affect hotel ratings. They empirically found that reviewers from countries with high power distance give lower online hotel ratings.

Recommender systems are widely used to suggest contacts, or activities on social media platforms (Wu *et al.*, 2014), and to improve targeted ads by the advertising industry (Buyya *et al.*, 2016). Practitioners can develop new marketing strategies by integrating users' current situations and future needs by offering contextually relevant socialized recommendations (Shen *et al.*, 2013). Lu *et al.* (2015) presented a comprehensive analysis of recommender systems applications reviewing up-to-date application developments and clustering them into eight main categories.

In accordance with the above, even from Ranganathan's (1931) "The Five Laws of Library science", the question of intelligent exploitation of new technologies is raised to improve user service for human needs and creating the future. Alvin Toffler in his book Future Shock (1971) set the theory that the information overload will lead to decision-making conflict. A comprehensive overview of information overload referring to too much information identified by the academic community was presented by Eppler and Mengis (2004): excessive communication overload (Meier, 1963), information overload of sensors (Lipowski, 1975), cognitive overload (Vollmann, 1991), information knowledge overload in medicine (Hunt and Newman, 1997), syndrome of information fatigue (Wurman, 2001). Also, there are studies that have confirmed and located other types of information overload within the professional service sector (Srinivasan, 2016) ranging from business consulting (Hansen and Haas, 2001) to management meetings (Grisé and Gallupe, 1999).

According to Teets, the collation of information has been going on for centuries. One of the first recorded attempts was in the middle of the third century BC when *Pinakes* organized a library catalog, by listing author's names in alphabetical order (Weinberger, 2007). Little seems to have changed in the meantime – information in library catalogs were still primarily organized for the physical world (Eliot and Rose, 2009). The significant innovation occurred in 2010 when the German National Library released a library catalog linking authors' works to others in the same fields (Svensson and Jahns, 2010). In 2011, the British Library published national library data as a linked data, describing the model of Things of Interest, where the book title linked to people, events, and places (Hodder, 2013; Teets and Goldner, 2013). In 2009, Tim Berners-Lee recommended the first step which was to set up data and information on the web in a form in which machines can naturally understand or convert into an understandable form that will lead to the end result of this process – Linked Data (Bizer *et al.*, 2009).

This is precisely the issue because modern data library systems should provide insights by utilizing various techniques, such as Big Data analytics of all type and sources of data; they should enable business analytics and real-time processing whilst increasing the capacity to deliver significant data and content of interest through the recommender system. Such system provides comprehensive logistics and an analytical platform with the fully featured tools and solutions to meet the needs of the most sophisticated and demands of modern, smart library systems.

Traditional and e-library recommender systems

Traditional library systems of educational institutions are usually configured as RDBMS, and they are still very wide-spread. As such, they lose the ability to process large logs, text, images, audio, video, sensor records and other complex types of data that arrive at a high rate from different sources. At present, the structured database systems that store the vast majority of organizational data are unsuitable for analytical processing. Difficulties also lie in data capture, sharing, and visualization (Ahrens *et al.*, 2011). Big Data applications and systems are built to respond to these emerging challenges. Using Big Data frameworks allows practitioners to make decisions based on evidence rather than on intuition. Traditional library systems often suffer in both, memory use and storage capacity (Bekkerman *et al.*, 2011). Valuable data in those systems that may carry important information, collected and stored at a high cost is ignored and finally deleted because of limited warehouse space (Worlton, 1971).

Now, when libraries are faced with the Tony Hays fourth paradigm of data-intensive science and overwhelming data sets from many different sources, it becomes more and more difficult for a library to provide enough space to store all the necessary information that is important for users, and submitted for its long-term keeping. (Hey *et al.*, 2009; Horstmann and Brase, 2016; Bhat, 2018). Furthermore, in the traditional RDBMS of the library, the recommendation is made by the librarians based on their experience, user's physical cards, and the relational database server logs. The suitability of such a recommendation is not knowledge-based and consumes both, a lot of time and other resources. Recommender systems are used in e-library applications to help users locate and select information and knowledge-based sources (Porcel and Herrera-Viedma, 2010). The hybrid-based recommender system Fab, part of the Stanford University Digital Library Project, which combines CB and CF recommendation algorithms, was presented by Balabanović and Shoham (1997). Mooney and Roy (2000) developed a book recommender system with a machine-learning algorithm for text categorization. A Naive Bayesian text classifier utilized information extraction to build features of books and user preferences to find the best matched books for the observed user. Later, a personalized e-library service called CYCLADES was proposed by Renda and Straccia (2005). CYCLADES provides recommendation algorithms which rely on personalized information of the organization and users' opinions in an integrated environment. Based on the research of Porcel *et al.* (2009), who developed a hybrid-based recommender to advise research resources in University Digital Libraries (UDL) to handle flexible information by means of linguistic labels by creation users' preferences relation, Serrano-Guerrero *et al.* (2011) presented a recommender engine which can incorporate GoogleWave technology in UDL.

All of the discussed e-library recommender systems are mostly using hybrid-based recommender approaches which combine CB and CF techniques (Lu *et al.*, 2015).

E-learning and Big Data recommender systems

Recommender systems have found applications in diverse domains such as e-learning and Big Data science. Since the early 2000s, e-learning recommender systems have been increasingly popular. Based on the development of traditional e-learning systems and more than fifteen years of research studies on this topic, many practical and applicable solutions of e-learning recommenders have been developed. This type of system aims to help students choose courses, and find learning materials that they are interested in. Zaíane (2002) proposed building a software agent that uses data mining techniques, based on association rule algorithm for constructing a model which represents the behavior of a user. A personalized e-learning material recommender system was proposed by Lu (2004). When a student's registration is obtained in the database, the system uses a computational model to identify user learning preferences which it combines with matching rules to generate a recommendation. Chen and Chao (2008) developed a system that augments traditional books with online discussion forums and learning communities. Based on their preferences, members receive messages from a web-based learning community which includes links to additional online resources. Romero *et al.* (2009) proposed an e-learning recommender system that utilizes Web usage mining to recommend links in a Web-based educational system. Further, a hybrid-based e-learning recommendation approach was developed by Capuano *et al.* (2014). This system prototype recommends learning goals generating recommendations through learning experiences and user needs. A hybrid-based approach with a three-step recommendation: mapping, utility estimation, and higher level learning goals. Extending web-based educational systems with personalized support through user centered designed recommendations was proposed by Santos *et al.* (2014). This study shows that the building of a personalized e-learning system is a process that needs to address students' needs throughout the e-learning life cycle. Recently, De Meo *et al.* (2017) formed e-learning classes by evaluating trust and skills of learners and proposed a model aimed at managing the formation and the evolution of e-learning classes based on the information available on online social networks.

Some of the relevant research studies that have linked Big Data recommenders with data analytics in other areas include (Khan *et al.*, 2017): analytics in business (Duan and Xiong, 2015), climate changes (Lu *et al.*, 2011), analytics of bank customers (Sun *et al.*, 2014), smart cities (Khan *et al.*, 2015), social media analytics (Burnap *et al.,* 2015), healthcare analytics (Raghupathi and Raghupathi, 2014), railway management system analytics (Thaduri *et al.*, 2015) and intelligent transportation (Chandio *et al.*, 2015). Also, Saboo *et al.* (2016) proposed a time-varying effects model for handling the complexities associated with Big Data analytics for resources (re)allocation in marketing strategies. He *et al.* (2014) presented a social recommender system based on Hadoop (SRSH) to generate recommendations of similar users and user communities for finding friends and content. Ismail and Al-Feel (2015) proposed a Hadoop-based recommendation system for research papers. In the same year, Yao *et al.* (2015) developed a Big Data-based Hadoop ecosystem for facilitating data processing in healthcare services and clinical research. Wang (2016) designed and implemented a network recommendation system based on the Hadoop platform. Yi *et al.* (2017) presented a library recommendation method based on association rules combined with an artificial bee colony algorithm with the aim of producing personalized booklists using historical borrowing records. A multimedia recommender system for online social networks named SOS was presented by Amato *et al.* (2017). Integration NoSQL and relational database into the Hadoop ecosystem was the project of Rodrigues *et al.* (2018) with the aim to implement an e-commerce prototype system to manage credit card transactions, involving large volumes of data by using different technologies.

The structure of each of these works is different. It is important to understand that one of the lesser-explored applications of Big Data analytics lies in the smart library recommenders of an educational institution. Moreover, the use of this synergistic approach for developing a Dig Data smart library platform for the implementation of diverse data sets and sources needs to be explored. The need for research in this field can be summarized as a lack of a comprehensive survey of the use of huge amounts of data from differential sources when creating smart library applications in a learning continuum that can benefit both, educational institution and users' needs.

Big Data smart library model

This paper presents a Big Data smart library model with the aim of building a system that can recommend personalized content to users with increased precision by analyzing users interests collected from multiple sources, as well as the characteristics of content from different types of data. This process enhances the quality of the recommendation scenario as treats the relationship between the libraries and education as inseparable from one another.

It can be said that there are three key strands which make this system critical for information companies which manipulate large amounts of data such as libraries. They are: switching to a scalable and elastic infrastructure, the complexity and diversity of available data, the power and value of combining different types of data.

The proposed model shown in Figure 1 contains the main components necessary for the creation of a smart library. As shown, data is collected into the smart library from multiple data sources including the LMS (Despotović-Zrakić *et al.*, 2012), the educational institution IS, social media networks (Hargittai, 2007), online bookstore server logs (Menascé *et al.*, 1999), and the IoT (Barnaghi *et al.*, 2012). The primary focus of the smart library shown in

Figure 1 is the integration of multiple sources of different data and providing personalized recommendations to the user. The goal of the new Big Data smart library is the realization of a more precise and efficient recommender system. The interaction between different data sources in generating personalized recommendations is presented in an innovative way with the accessible system throughout data integration which achieves the most interoperability.

The smart library connects and combines data from the following sources: information that students exchange between themselves and teachers via the news forum on the Moodle platform, including selected student courses during the learning cycle; the data set from the IS of the educational institution; information collected from social media networks; the server log files from the University online bookstore that contains personalized customer information; information collected from IoT sensors, i.e., location of the printed edition of a book in the library and its usage. This data is gathered, processed and analyzed in the Big Data ecosystem which has been selected for the realization and visualization of the final content – personalized content-aware recommendations to the user based on his/her interests.

Figure 2 is a flowchart of the realized Big Data ecosystem for the smart library.

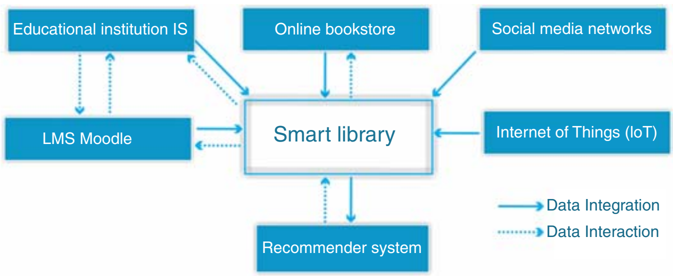


Figure 1.
Smart library block diagram

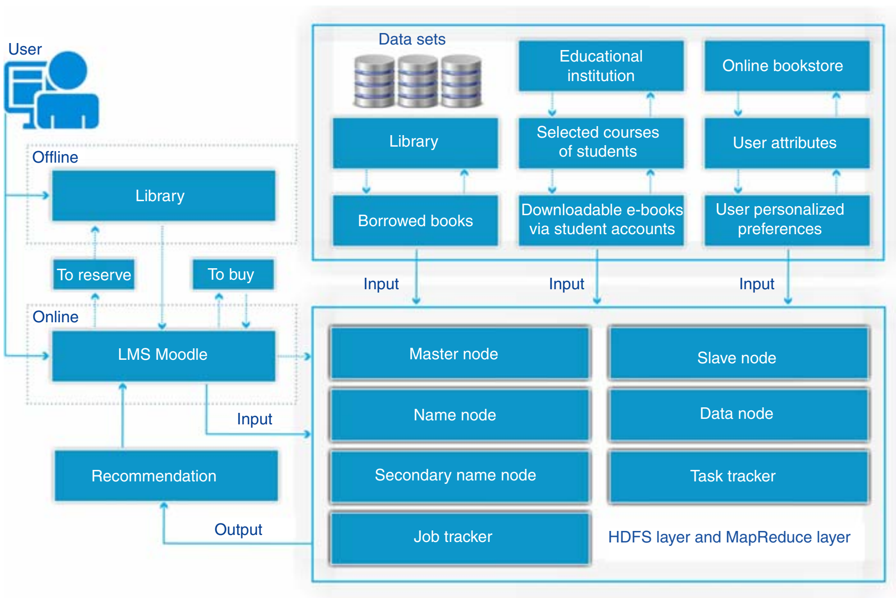


Figure 2.
Flowchart of the realized ecosystem for the smart library

Data sets from various sources are loaded and integrated into the HCatalog. These different data sets include: borrowed books from the University library; downloadable e-books via student accounts through the IS of the educational institution; server logs from the university online bookstore that contain information about the users' personal preferences; and the student's preferences in the current school year on the LMS platform.

With every new access of a registered user on the LMS platform, the Hadoop ecosystem analyzes the following: information about books attributes from the user's logged book usage in the library; data from the IS of the educational institution about the selected courses of the student in the current school year; the purchase history and user attributes from the University online bookstore logs; performs CF on multiple data; and at the completion of the processing of given data, the user receives a recommendation of greater interest and precision with two available choices:

(1) Choice 1: to reserve the recommended books in the library of the educational institution in which case the librarian performs a reservation for the user under his ID for a specified time period for borrowing.

(2) Choice 2: to buy the recommended books from the University online bookstore.

Data sets description

A data set of the educational institution for three years of undergraduate studies contains 160 courses, seven study programs and ~2,250 students from which approximately available books to download from a student service per student in the current school year is 5-15, depending on selected subjects and available literature, with the overall of over 10 million records.

The library data set contains 470,571 users and 3,955 book titles. The University online bookstore logfile that was used contains ~3,900 evaluations, ~450 users and ~1,495 items that were taken for the calculation. For these data sets and their analysis, an important factor is coverage (Sarwar *et al.*, 2000). Coverage is a percentage of the total number of items for which the system will generate recommendations. The basic measurement of coverage is the percentage of available items. It is the percentage of all users of the system for which the forecast is requested. The common characteristics of systems that can reduce coverage are small dimensions of similarity of users and their sampling. Logfile has a high coverage of approximately 95 percent. Maximum coverage is not provided for the following reasons: there are certain items without evaluation (purchased or rated); small number of users who have evaluated a particular item; and low similarity of a particular user in the system.

Each student is required to apply to a course on the LMS Moodle platform for the subjects he has chosen in the current school year, which may be between 10 and 12 courses. The LMS Moodle platform contains the following information: three years of undergraduate studies (seven study programs), one year of specialist studies (six study programs) and two years of master studies (three study programs) – which is a total of 440 courses, ~3,750 students, ~900 teachers, and over 1,500 records per day.

Implementation and results

For the pilot implementation, the chosen platform was open-source Apache Hadoop Hortonworks. Apache Hadoop is a widely adopted and one of the most well-established Big Data software platforms that support distributed data-intensive application and the MapReduce computational paradigm that allows parallel processing of a huge volume of heterogeneous data. MapReduce and Hadoop are considered to be the most effective and efficient framework for Big Data management (Khan *et al.*, 2017).

In the implemented ecosystem, data are stored in HDFS, which provides scalable and fault-tolerant storage. HDFS detects and compensates cluster errors, splitting incoming files into blocks, and storing them redundantly over clusters. The files are divided into the blocks (64MB or 128MB each), and each block is copied on more than one node. A projected system with replica possibilities allows fault tolerance – where the loss of one node will not destroy the file, and performance – where any block can be read from one or more nodes while improving the data flow. HDFS in this model of the library ecosystem provides data availability by continuously monitoring the nodes in the cluster and the blocks they manage. The individual blocks are subject to checks and controls. When a block has been read, correctness is determined (whether the recorded value is correct). If the block is damaged, it has been replaced with one of its replicas from another cluster clone (Olson, 2010; White, 2012). Parallel data processing is executed by the MapReduce programming model. MapReduce provides large data sets to be shared in clusters for parallel processing. The master node assigns tasks to the slave node and then collects the results. The model thus defined has two main steps: map – the job distribution and reduce – collecting results (Dean and Ghemawat, 2008).

The Hadoop technology stack with modules of importance for the creation of a Big Data smart library is shown in Figure 3. MapReduce programming in the proposed library ecosystem is performed by the Hive module that allows for the execution of queries over large data sets and provides a mechanism for data structure projecting. The prominent feature of this layer is a structure subjected to parallelization which enables management of the large data sets in the ecosystem (Thusoo *et al.*, 2009; George, 2011). The specified data sets were loaded into the HCatalog, ready for further processing on the management layer of

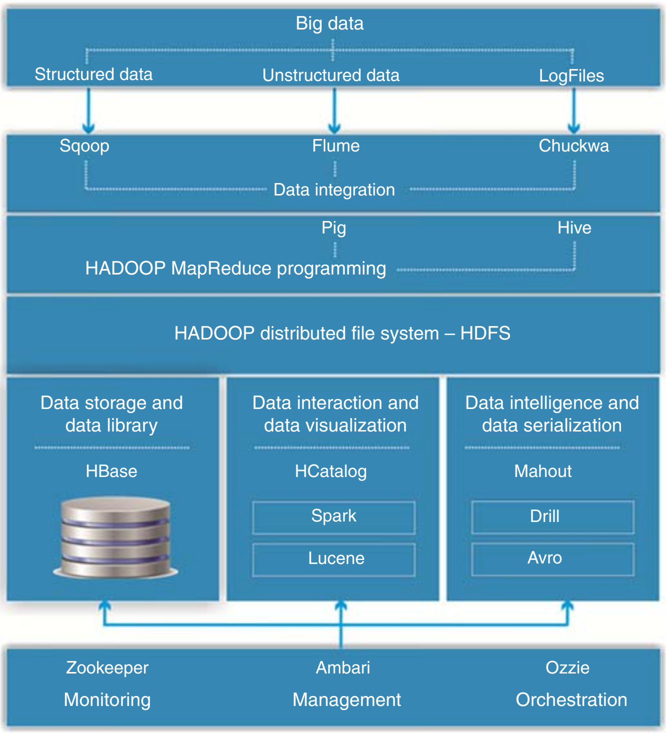


Figure 3.
Hadoop technology stack for the proposed smart library model

the ecosystem. The data is later distributed and transferred to the HBase module which enables searching, downloading and analyzing.

The query performed on the library data (shown below) aims to display users who have borrowed specific books in a given time period in 2017, and to prepare the data set for further processing (Note: The query was translated into English for easier understanding. At the Ambari module in Figure 4, the query is shown in its original form):

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”.

The search and selection of the most appropriate items were performed according to the following six steps:

(1) Step 1: the results of the executed query show book titles and the ID of a user who has borrowed the most books in the specified date range (shown in Table I).

(2) Step 2: monitoring and administration in the ecosystem are executed by the Ambari module which enables installation, management, and monitoring of the Hadoop services in the cluster. The integrated module: coordinates distributed applications, synchronizes and centralizes services in the cluster, and coordinates and monitors the workflow of mutually independent Hadoop jobs (White, 2012). The result of the executed query from Ambari module is shown in Figure 4.

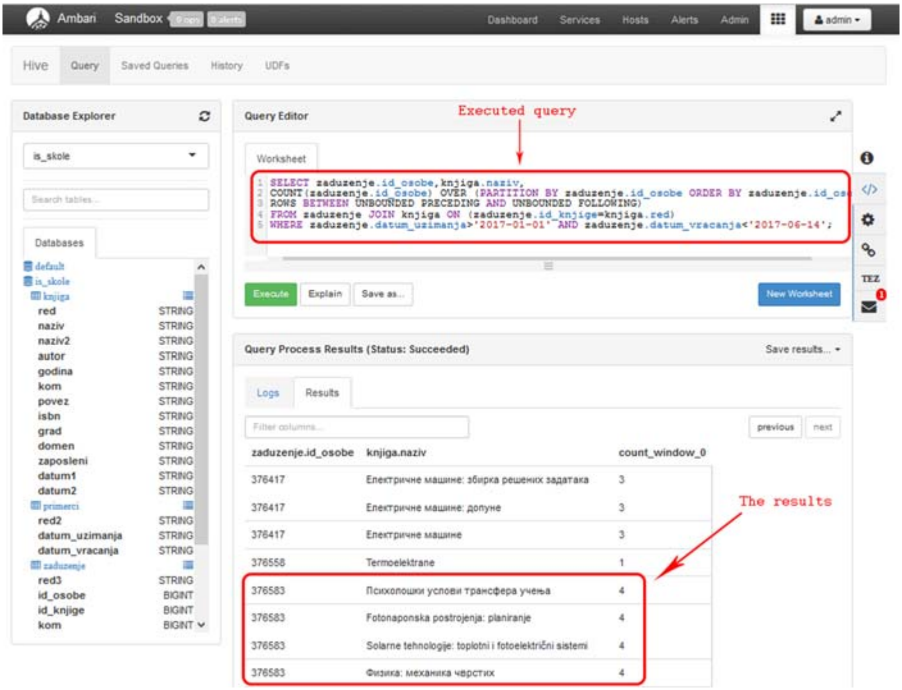


Figure 4.
Ambari – results of the executed query

(3) Step 3: in the implemented ecosystem, at the external and internal sources (the IS of the educational institution, the University library, the University online bookstore, and the LMS Moodle platform) the user ID is linked to the e-mail account.

(4) Step 4: on the online student services of the IS of the educational institution, students could, by using their accounts, download e-books on the basis of selected subjects for the current and previous school years. The available e-books for download for the current school year per student could be in the range from 5 to 15.

(5) Step 5: the recommender system which is integrated into the University online bookstore uses item-based CF (Linden *et al.*, 2003). At any new access by the user to the online bookstore, the system analyzes the users' purchase history, attributes, and personalized preferences to generate a list of recommendations and content of interest. The item-based CF compares each purchase made by users, ranks the items by similarity, and by combining the similar products generates a recommendation (Simović, 2014). To every user who has purchased or rated items from the University online bookstore, the system recommends three books.

(6) Step 6: at a user's new logon the LMS Moodle platform, the proposed Big Data recommendation system processes the data from the University library, the IS of the educational institution, the online bookstore server logs, and the LMS platform to generate a recommendation. The application is developed using the Moodle API. The algorithm calculates similarity to determine which items are most suitable for the user based on all the items in the four separate systems. The way in which the algorithm calculates the similarity of items with all others in HCatalog is shown in the following pseudo-code.

```
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
```

Figure 5 shows a recommendation to a user with the same ID that the Hadoop ecosystem generates on the LMS Moodle platform using four independent sources and four integrated

Table I.
User ID and book titles

User ID	Book titles (books borrowed from the library)
376583	*Психолошки услови трансфера учења* (psychological conditions of learning transfer)
376583	*Fotonaponska postrojenja: planiranje* (photovoltaic plants: planning)
376583	*Solarne tehnologije: toplotni i fotoelektrični sistemi* (solar technologies: thermal and photoelectric systems)
376583	*Физика: механика чврстих* (physics: solid mechanics)

Figure 5.
Hadoop recommender on the LMS moodle platform

data sets. The user receives the following choices: to reserve the recommended books in the library of the educational institution; or to buy the recommended books from the University online bookstore. The three book titles that the Big Data ecosystem generates for the user with the same ID are shown in Table II.

Evaluation of the system

The evaluation of the system was conducted in two ways. For the first evaluation, Big Data analysis of the system was used to determine the most recommended books in the Hadoop ecosystem in order to discover whether those books were borrowed more often from the University library compared to the same period of the previous year.

For the second evaluation of the system, an online questionnaire has been chosen as an activity for the evaluation of the implemented prototype ecosystem. The questionnaire was used to determine whether the Hadoop-based recommendation with multiple sources is more appropriate for the users (based on their impressions) than the University online bookstore recommendation system.

Big Data analysis of the system

In the first evaluation, Big Data analysis was used to determine the following: which books were recommended to users most often during the evaluation of the system prototype; and how many times were the recommended books borrowed from the University library in 2016 in comparison to 2017. The aim of the analysis was to evaluate the potential of the Hadoop-based recommendation ecosystem to contribute an improvement of the business performance of the organization and to an increase of the users' usage and trust in the system.

User ID	Book titles (recommended on LMS Moodle based on Hadoop ecosystem)
376583	Увод у проналажење информација на вебу (introduction to finding information on the web)
376583	Одрживи развој (sustainable development)
376583	Електронско банкарство (electronic banking)

Table II.
LMS Moodle recommendation based on Hadoop

The following query shows which books were most often recommended to users on the LMS Moodle platform based on the Hadoop ecosystem:

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).

The results of the executed query from the Ambari module are shown in Figure A1 and Table III.

Further, the two queries were executed over two separate time periods (2016 and 2017), which were then compared. The first query determines how often recommended books were borrowed from the University library in a given time period for 2016 before the prototype of the Big Data ecosystem for the smart library was activated:

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 ("1,016", "844", "1,002", "138", "1,463", "1,497", "103", "118", "571", "1,916", "2,113", "2,275", "7", "264")

AND recommender.date_taken > "2016-1-1" AND recommender.date_return < "2017-1-1".

The results of the executed query are shown in Figure A2.

The second query determines how many times the books were borrowed from the University library in a given time period in 2017 after the activation of the prototype of the Big Data ecosystem for the smart library:

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)

ID of the books	Book titles	Number of times
1016	*Osnovi elektronike i telekomunikacija (Basics of electronics and telecommunications)*	3
844	*Хидраулика: увод са примерима управљања (Hydraulics: introduction to control, examples)*	2
1002	*Моторна возила I: општи и теоријски део (Motor vehicles I: general and theoretical part)*	2
138	*Електричне машине: за трећи разред електротехничке школе (Electrical machines: for the third grade school of electronics)*	3
1463	*Komutatorni motori (Commutate motors)*	3
1497	*Termoelektrane (Thermal power plants)*	3
103	*Физика: механика чврстих (Physics: solid mechanics)*	6
118	*Трансформатори (Transformers)*	6
571	*Психолошки услови трансфера учења (Psychological conditions of learning transfer)*	6
1916	*Energetski transformatori i generatori (Energy transformers and generators)*	6
2113	*Фотонапонска постројења: планирање (Photovoltaic plants: planning)*	6
2275	*Solarne tehnologije: toplotni i fotoelektrični sistemi (Solar technologies: thermal and photoelectric systems)*	6
7	*Основи рачунарске технике (Basics of computer technology)*	2
264	*Математички приручник (Mathematics manual)*	2

Table III. The most recommended books to users

FROM recommender JOIN book ON (recommender.book_id = book.count)

WHERE recommender.book_id IN ("1,016", "844", "1,002", "138", "1,463", "1,497", "103", "118", "571", "1,916", "2,113", "2,275", "7", "264")

AND recommender.date_taken > "2017-1-1" AND recommender.date_return < "2017-12-1";

The results of the executed query are shown in Figure A3.

Comparative overview of recommended books borrowed in 2016 and 2017 is shown in Table IV.

As shown in Figure 6, in 2016, the books that were taken for the evaluation of the system were borrowed 23 times, compared to 2017 when the same books were borrowed 62 times which is approximately 269.5 percent higher.

Student's impressions

The second evaluation of the system consisted of an online questionnaire as the method for the evaluation of the implemented prototype ecosystem. The questionnaire was integrated into the LMS Moodle at the School of Electrical and Computer Engineering of Applied Studies, University of Belgrade and used to determine whether the Hadoop-based

Recommended books (IDs)	Book titles (the most recommended books to users on Hadoop ecosystem)	Borrowing (2016) (number of times)	Borrowing (2017) (number of times)
1016	*Osnovi elektronike i telekomunikacija*	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	*Komutatorni motori*	1	1
1497	*Termoelektrane*	4	4
103	*Физика: механика чврстих*	1	1
118	*Трансформатори*	1	4
571	*Психолошки услови трансфера учења*	1	1
1916	*Energetski transformatori i generatori*	1	25
2113	*Фотонапонска постројења: планирање*	1	9
2275	*Solarne tehnologije: toplotni i fotoelektrični sistemi*	4	2
7	*Основи рачунарске технике*	4	4
264	*Математички приручник*	1	1

Table IV.
Comparative overview

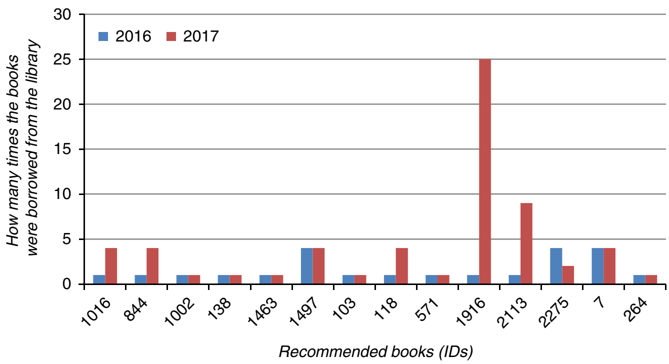


Figure 6.
Number of times that the recommended books have been borrowed

recommendation with multiple sources is more appropriate for the users (based on their impressions) than the University online bookstore recommendation system.

The sample was 220 of students in the third year of undergraduate studies in the area of E-business and ISs during the winter semester of the 2016/17 school year. All students agreed to take part in the research. The questionnaire included ten questions. Each question was supported by Moodle: answers from multiple choices and true-false answers. Two important questions were based on the five-point Likert scale. The online questionnaire is shown in Table V.

Analysis of results

The first consideration is a Big Data analysis that illustrates the potential benefits of the application of a Big Data smart library with the integration of multiple sources of differential data in the context of meeting user needs and improving the economic benefits for the organization with the overall satisfaction of its services, through IT project management thereby promoting mutual understanding and the development of long-term relationships (Liu *et al.*, 2017). That is confirmed by the results which are showed an increase of approximately 269.5 percent in the number of books that were borrowed from the University library from the most recommended ones on the LMS Moodle platform based on Hadoop during the prototype system testing.

The second considerations dealt with the registration and purchases from the University online bookstore. The results showed that 40 percent of the total number of students, who took part in the online questionnaire, did not register and make a purchase at the online bookstore. Without the registration and purchases, the system was unable to determine the user's preferences and therefore could not create a model of the user and generate a recommendation because of a cold start problem (Schein *et al.*, 2002). This group of students was excluded from the final results of the Hadoop-based recommendation on the LMS Moodle platform. Other students, which had access to e-books for download from their online student services from the IS of the educational institution, and which have made purchases at the University online bookstore, and also borrowed books from the University library, met the requirements for the evaluation survey.

No.	Question
1	Do you borrow books from the University library? (a) Yes; (b) No
2	During the 2016/17 school year, how many books did you borrow? Depending on the previous question: (a) < 5; (b) $\geqslant 5$
3	On the LMS Moodle, do you get a book recommendation? (a) Yes; (b) No
4	Please evaluate the received recommendation on the LMS Moodle platform based on Hadoop recommender (If the answer to the question No. 3 is yes): (a) Completely inadequate; (b) Somewhat inadequate; (c) Neutral; (d) Somewhat adequate; (e) 1Completely adequate
5	Based on the selected subjects in the school year 2016/17, in the information system of the educational institution, on your online student service, do you have available e-books to download? (a) Yes; (b) No
6	How many e-books are available for download through your online student service? Depending on the previous question: (a) < 5; (b) $\geqslant 5$
7	Have you registered and made a purchase on the University online bookstore? (a) Yes; (b) No
8	How many books did you buy on the University online bookstore? Depending on the previous question: (a) < 5; (b) $\geqslant 5$
9	When you sign into the University online bookstore, do you get a book recommendation? (a) Yes; (b) No
10	Please evaluate the received recommendation on the University online bookstore based on collaborative filtering (CF) (If the answer to the question No. 7 is yes): (a) Completely inadequate; (b) Somewhat inadequate; (c) Neutral; (d) Somewhat adequate; (e) Completely adequate

Table V.
An online questionnaire for gathering data from students based on their impressions

The results of the questionnaire data are shown in Table VI. They show that – on the basis of the examined participants and their impressions, with the results of the survey analysis ($\overline{x}$ – mean grade, from 1 to 5 and δ – standard deviation) and by the integration of multiple sources of different data – the system generates more appropriate recommendations, although there are small differences. In all, 24.6 percent of students rated the Hadoop recommendation completely adequate with a mean grade of 3.38 and a standard deviation of 1.25, while 21.1 percent of students have rated the recommendation of the University online bookstore completely adequate with a mean grade of 3.17 and a standard deviation of 1.28. Answers to questions related to the recommender system show that students' slightly favor the Big Data recommendation that was integrated into the LMS Moodle platform and the Hadoop framework.

Discussion

According to the Lapkin report in Gartner (2012) "Big Data is not just about MapReduce and Hadoop. Although many organizations consider these distributed processing technologies to be the only relevant Big Data technology, there are alternatives. In addition, many organizations are using these technologies for more traditional use cases, such as preprocessing and the staging of information to be loaded into a data warehouse." Such alternatives for parallel database systems for parallel processing and data analysis (e.g. Vertica, Teradata, Netezza, SQL Server, Greenplum, ParAccel) are expensive, difficult for administration, and have deficiency in fault tolerance and processing speed for long-running queries (Pavlo *et al.*, 2009; Sakr, 2016). In practice, Hadoop as an open-source project has achieved great success, with increasing momentum in research and development in educational and business domains. This technology, with modules of importance and their combination, has enabled even small companies to collect and analyze Big Data in order to gain a competitive advantage. Hadoop as open-source software provides a tool to process vast amounts of data easily and cost-effectively (Sakr, 2016). On the flip side, organizations need a data scientist to establish a workflow and fully utilize the advantages of Big Data analytics. Further, poorly identified input data will result in poor quality results regardless of the reliability of the Big Data solution.

The major challenge in building a smart library is the analysis of a large amount of data that can be carried out in parallel processing through the Hadoop ecosystem. Proper analysis of the multiple data of differential sources integrated into the smart library is not only useful for increasing the efficiency and economic benefit of the organization but is also effective in additionally motivating users by providing hedonistic values.

The obtained results show that the proposed system positively influences the performance parameters of the library by increasing the number of borrowed books and the perception of students, which indicate that such a system can be applied as an integral part

Question	Completely adequate (%)	Somewhat adequate (%)	Neutral (%)	Somewhat inadequate (%)	Completely inadequate (%)	$\overline{x}$	δ
Please evaluate the received recommendation on the LMS Moodle platform based on Hadoop recommender	24.6	22.8	26.3	19.3	7.0	3.38	1.25
Please evaluate the received recommendation on an the university online bookstore based on collaborative filtering (CF)	21.1	19.3	24.6	26.3	8.8	3.17	1.28

Table VI. Descriptive statistics of students ($n = 132$) based on their impressions with means and standard deviations related to questions No. 4 and No. 10

of learning processes making the approach particularly suitable for application in educational institutions. By using the proposed data analysis system, librarians can achieve Big Data smart library operation for the coming generations.

Conclusion

The objective of this research is focused on the proposal of a solution for managing large differential data from multiple sources in smart libraries based on the Hadoop ecosystem. With integrating recommender systems to the smart library in the Big Data environment, satisfaction value for the users and the unique features for library management are proposed. The limitations of this study and the implications for research and practical use are presented in the following sections.

Limitations

Before the discussion of the possible limitations, it should be noted that most business systems are aware of the fact that the application of Big Data technology is inevitable, but are still trying to delay its implementation. Clearly, Big Data analytics requires the engagement of Data Scientists (Davenport and Patil, 2012) whose role is now significantly changing in library business systems. In order to achieve a competitive advantage in the Big Data environment, libraries need to implement a Big Data framework and the appropriate modules and tools that will increase customer satisfaction and long-term trust (Ratledge and Sproles, 2017).

This paper presents a practical application of a Big Data recommender system and the first step of creating fully operational smart library system. The loading and processing of four data sources were carried out. The obtained results were visualized and evaluated. The implemented storage architecture has limitations and disadvantages when it comes to large-scale distributed systems. Required applications for running on highly scalable clusters of computers are server-side flow, data access optimization (Ishii and de Mello, 2012), and improved performance for data replication, distribution, migration, and access to parallelism (Philip Chen and Zhang, 2014). In order to improve performance and exploit the full potential of the proposed Big Data model in the smart library ecosystem, future integration with multiple (i.e. IoT and social media network) sources is needed. Also, data sparsity (Guo *et al.*, 2014) remains an issue.

Implications for research

The infrastructure of the proposed ecosystem applicable in library management at the data set layer has the ability to manage both, relational and non-relational structures, enabling decision support systems and it can manage data in real-time using streaming tools and modules for interoperability of internal and external data sources. It also provides analytical applications for prediction analysis and interactive research of pooled data from the different sources. The platform also needs to support Data Scientists and researchers in managing real-time decision making.

Implications for practice

The rapid spread and expansion of data and different sources has created new approaches to their processing and a new channel for ubiquitous information access in libraries. The Hadoop framework is a powerful tool for solving large data management problems in smart environments. Big Data technology has a fundamental role in information management and in the context of knowledge discovery and library operations improvements.

The proposed ecosystem has the capability to improve library operations for several reasons: ability to collect data from multiple different sources, reliability and fault tolerance in comparison to traditional data management systems, ability to store personal preferences and user attributes important for further predictions, analyzes content to find similarity from multiple sources, addresses the issue in the context of simultaneous access to multi-tier data from multiple sources, scales resources that deal with managing a larger amount of data, analyzes and processes data using modern Big Data technology, sharing of insights throughout of IT systems of the organization.

The practical implications of this work also contribute a hedonic value (Shen *et al.*, 2013) obtained using the recommendation system in smart libraries in a large data environment that will not only provide users useful and highly personalized contents but will also build long-term customer trust by generating a list of recommendations with greater precision.

Further work

The implementation of a smart library system with integrated IoT technologies and social media data, requires further research of the realized Big Data ecosystem that focuses on the following issues: Integration of data from social media platforms for real-time analytics; integration with the IS of the educational institution for developing and enabling a fully operational smart library system; IoT integration with distributed sensors in the library and how they generate data that is further processed in the proposed Big Data model in the Hadoop ecosystem in the most efficient way.

References

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.

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.

Amato, F., Moscato, V., Picariello, A. and Piccialli, F. (2017), "SOS: a multimedia recommender system for online social networks", *Future Generation Computer Systems*, available at: https://doi.org/10.1016/j.future.2017.04.028 (accessed April 23, 2017).

Balabanović, M. and Shoham, Y. (1997), "Fab: content-based, collaborative recommendation", *Communications of the ACM*, Vol. 40 No. 3, pp. 66-72.

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.

Bekkerman, R., Bilenko, M. and Langford, J. (2011), *Scaling up Machine Learning: Parallel and Distributed Approaches*, Cambridge University Press, New York, NY.

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.

Beyer, M.A. and Laney, D. (2012), "The importance of 'big data': a definition", Gartner, June 21, available at: www.gartner.com/id=2057415 (accessed May 15, 2017).

Bhat, W.A. (2018), "Long-term preservation of big data: prospects of current storage technologies in digital libraries", *Library Hi Tech*, available at: https://doi.org/10.1108/LHT-06-2017-0117

Bizer, C., Heath, T. and Berners-Lee, T. (2009), "Linked data-the story so far", *International Journal on Semantic Web and Information Systems*, Vol. 5 No. 3, pp. 1-22.

Boyd, D. and Crawford, K. (2012), "Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon", *Information, Communication, & Society*, Vol. 15 No. 5, pp. 662-679.

Buckland, M.K. (2017), "Library technology in the next 20 years", *Library Hi Tech*, Vol. 35 No. 1, pp. 5-10.

Burnap, P., Rana, O., Williams, M., Housley, W., Edwards, A., Morgan, J. and Conejero, J. (2015), "COSMOS: Towards an integrated and scalable service for analyzing social media on demand", *International Journal of Parallel, Emergent and Distributed Systems*, Vol. 30 No. 2, pp. 80-100.

Buyya, R., Calheiros, R.N. and Dastjerdi, A.V. (2016), *Big Data Principles and Paradigms*, Elsevier Inc., Morgan Kaufmann, Cambridge, MA.

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, available at: https://doi.org/10.1016/j.chb.2013.07.036

Chambers, S. (Ed.) (2013), *Catalogue 2.0: The Future of the Library Catalogue*, Facet Publishing, London.

Chandio, A.A., Tziritas, N. and Xu, C.Z. (2015), "Big-data processing techniques and their challenges in transport domain", *ZTE Communications*, Vol. 1 No. 10.

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.

Cisco (2017), "The zettabyte era: trends and analysis", Cisco, pp. 1-32, available at: www.cisco.com (accessed May 15, 2017).

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.

Davenport, T.H. and Patil, D.J. (2012), "Data scientist", *Harvard Business Review*, Vol. 90 No. 5, pp. 70-76.

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, available at: https://doi.org/10.1016/j.ins.2017.04.002

Dean, J. and Ghemawat, S. (2008), "MapReduce: simplified data processing on large clusters", *Communications of the ACM*, Vol. 51 No. 1, pp. 107-113.

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.

Duan, L. and Xiong, Y. (2015), "Big data analytics and business analytics", *Journal of Management Analytics*, Vol. 2 No. 1, pp. 1-21.

Edjlali, R., Feinberg, D., Beyer, M.A. and Adrian, M. (2012), "The state of data warehousing in 2012", Gartner, available at: www.gartner.com/id=1922714 (accessed May 10, 2017).

Eliot, S. and Rose, J. (2009), *A Companion to the History of the Book*, Vol. 98, John Wiley & Sons, Oxford.

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.

Erl, T., Khattak, W. and Buhler, P. (2016), *Big Data Fundamentals: Concepts, Drivers & Techniques*, Prentice Hall Press, New Jersey, NJ.

Feisel, L.D. and Rosa, A.J. (2005), "The role of the laboratory in undergraduate engineering education", *Journal of Engineering Education*, Vol. 94 No. 1, pp. 121-130.

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 No. 7209, pp. 47-50.

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, available at: https://doi.org/10.1016/j.tourman.2017.10.007

George, L. (2011), *HBase: The Definitive Guide: Random Access to Your Planet-Size Data*, O'Reilly Media, Inc., Sebastopol, CA.

Grisé, 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.

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.

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.

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.

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.

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.

Hey, T., Tansley, S. and Tolle, K.M. (2009), *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Vol. 1, Microsoft research, Redmond, WA.

Hodder, I. (2013), *The Meanings of Things: Material Culture and Symbolic Expression*, Routledge, Abingdon.

Horstmann, W. and Brase, J. (2016), "Libraries and data – paradigm shifts and challenges", *Bibliothek Forschung und Praxis*, Vol. 40 No. 2, pp. 273-277.

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.

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.

Ismail, A.S. and Al-Feel, H. (2015), "Digital library recommender system on hadoop", *Network Cloud Computing and Applications (NCCA), 2015 IEEE Fourth Symposium*, June, pp. 111-114.

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.

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, p. 2.

Laney, D. (2001), "3D data management: controlling data volume, velocity and variety", *META Group Research Note*, Vol. 6 No. 70.

Lapkin, A. (2012), *Hype Cycle for Big Data*, Gartner Group, Stamford, CT, available at: www.gartner.com/doc/2100215/hype-cycle-big-data- (accessed August 15, 2017).

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.

Lin, Z. (2014), "An empirical investigation of user and system recommendations in e-commerce", *Decision Support Systems*, Vol. 68, pp. 111-124, available at: https://doi.org/10.1016/j.dss.2014.10.003

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.

Lipowski, Z.J. (1975), "Sensory and information inputs overload: behavioral effects", *Comprehensive Psychiatry*, Vol. 16 No. 3, pp. 199-221.

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.

Lu, J. (2004), "A personalized e-learning material recommender system", *2nd International Conference on Information Technology and Applications*, Harbin, January 8-11.

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, available at: https://doi.org/10.1016/j.dss.2015.03.008

Lu, S., Li, R.M., Tjhi, W.C., Lee, K.K., Wang, L., Li, X. and Ma, D. (2011, November), "A framework for cloud-based large-scale data analytics and visualization: case study on multiscale climate data", *Cloud Computing Technology and Science (CloudCom), 2011 IEEE 3rd International Conference*, pp. 618-622.

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.

Meier, R.L. (1963), "Communications overload: proposals from the study of a university library", *Administrative Science Quarterly*, Vol. 7 No. 4, pp. 521-544.

Menascé, D.A., Almeida, V.A., Fonseca, R. and Mendes, M.A. (1999), "A methodology for workload characterization of e-commerce sites", *Proceedings of the 1st ACM Conference on Electronic Commerce*, pp. 119-128.

Mooney, R.J. and Roy, L. (2000, June), "Content-based book recommending using learning for text categorization", *Proceedings of the fifth ACM Conference on Digital Libraries*, pp. 195-204.

Olson, M. (2010), "Hadoop: scalable, flexible data storage and analysis", *IQT Quart*, Vol. 1 No. 3, pp. 14-18.

Panetta, K. (2016), "Gartner's top 10 strategic technology trends for 2017", available at: www.gartner.com/smarterwithgartner/gartners-top-10-technology-trends-2017/ (accessed May 10, 2017).

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", *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data, ACM*, pp. 165-178.

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, available at: https://doi.org/10.1016/j.ins.2014.01.015

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.

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.

Raghupathi, W. and Raghupathi, V. (2014), "Big data analytics in healthcare: promise and potential", *Health Information Science and Systems*, Vol. 2 No. 1, pp. 3-10.

Ranganathan, S.R. (1931), *The Five Laws of Library Science*, Madras Library Association and Edward Goldston, Madras and London.

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.

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.

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", *Information Technology-New Generations*, Springer, Cham, pp. 443-451.

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.

Rubin, R.E. (2017), *Foundations of Library and Information Science*, American Library Association, New York, NY.

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. 911-940.

Sakr, S. (2016), *Big Data 2.0 Processing Systems: A Survey*, Springer International Publishing, New York, NY.

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, available at: https://doi.org/10.1016/j.scico.2013.12.004

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, Minneapolis, MN.

Schein, A.I., Popescul, A., Ungar, L.H. and Pennock, D.M. (2002), "Methods and metrics for cold-start recommendations", *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM*, pp. 253-260.

Shen, X.L., 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.

Showers, B. (2014), "Developing a shared analytics service for academic libraries", *Insights*, Vol. 27 No. 2.

Showers, B. (Ed.) (2015), *Library Analytics and Metrics: Using Data to Drive Decisions and Services*, Facet Publishing, London.

Simović, A. (2014), "Sistemi preporuke u e-trgovini", *Sinteza 2014-Impact of the Internet on Business Activities in Serbia and Worldwide*, pp. 846-852, doi: 10.15308/sinteza-2014-846-852.

Srinivasan, V. (2016), *The Intelligent Enterprise in the Era of Big Data*, John Wiley & Sons, Oxford.

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 Nos 5/6, pp. 4-1.

Svensson, L.G. and Jahns, Y. (2010), "PDF, CSS, RSS and other acronyms: redefining the bibliographic services of the German national library", *World Library and Information Congress: 76th IFLA General Conference and Assembly*, pp. 10-15.

Teets, M. and Goldner, M. (2013), "Libraries' role in curating and exposing big data", *Future Internet*, Vol. 5 No. 3, pp. 429-438.

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, available at: https://doi.org/10.1016/j.procs.2015.07.323

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.

Toffler, A. (1971), *Future Shock*, Bantam, New York, NY.

Vollmann, T.E. (1991), "Cutting the Gordian knot of misguided performance measurement", *Industrial Management & Data Systems*, Vol. 91 No. 1, pp. 24-26.

Wang, Q. (2016, August), "Design and implementation of recommender system based on Hadoop", *Software Engineering and Service Science (ICSESS), 2016 7th IEEE International Conference, IEEE*, pp. 295-299.

Weinberger, D. (2007), *Everything is Miscellaneous: The Power of the New Digital Disorder*, Macmillan, Basingstoke.

White, T. (2012), *Hadoop: The Definitive Guide*, O'Reilly Media, Inc., Sebastopol, CA.

Worlton, W. (1971), "Bulk storage requirements in large-scale scientific calculations", *IEEE Transactions on Magnetics*, Vol. 7 No. 4, pp. 830-833.

Wu, L., Shah, S., Choi, S., Tiwari, M. and Posse, C. (2014), "The Browsemaps: collaborative filtering at LinkedIn", In *RSWeb@ RecSys*.

Wurman, R.S. (2001), *Information Anxiety*, WUR, CIMMYT, New York, NY.

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.

Yi, K., Chen, T. and Cong, G. (2017), "Library personalized recommendation service method based on improved association rules", *Library Hi Tech*, available at: https://doi.org/10.1108/LHT-06-2017-0120

Zaíane, O.R. (2002, December), "Building a recommender agent for e-learning systems", *Computers in Education, 2002. Proceedings. International Conference, IEEE*, pp. 55-59.

Zikopoulos, P.C., Eaton, C., deRoos, D., Deutsch, T. and Lapis, G. (2012), *Understanding Big Data*, McGraw Hill, New York, NY.

Further reading

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.

Appendix 1

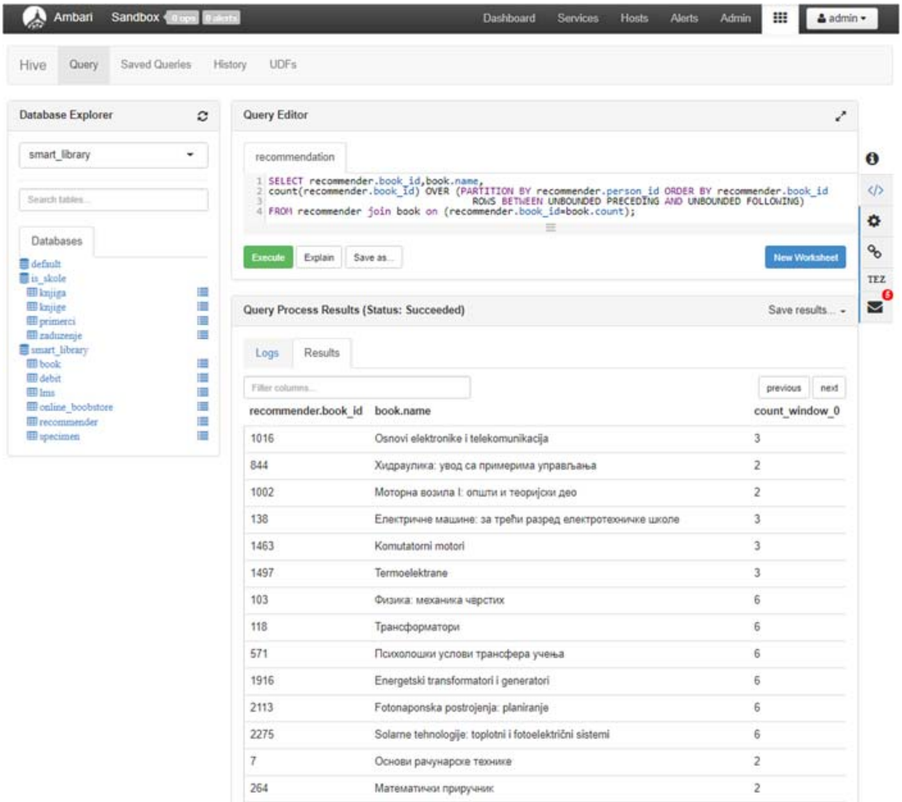
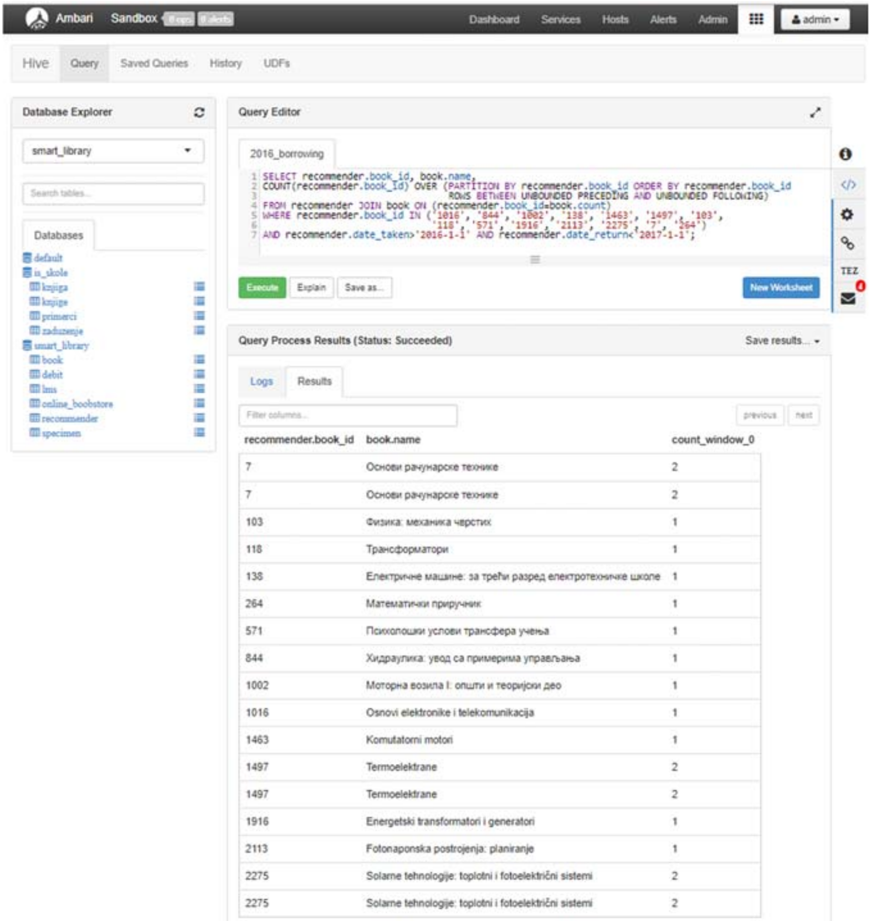


Figure A1.
Ambari module and the results of the executed query in order to determine the most recommended books in the Hadoop ecosystem

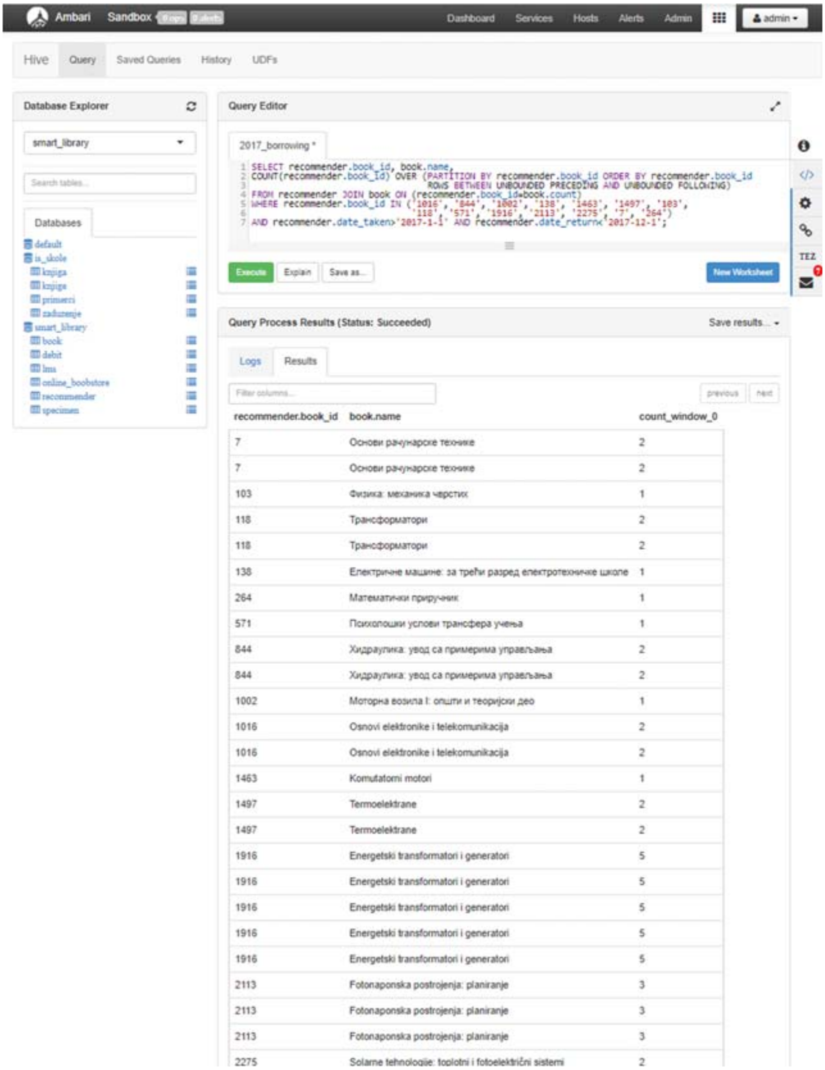
Appendix 2



Notes: Number of times that the recommended books have been borrowed from the University library in 2016. Before the activation of a Big Data ecosystem for the smart library

Figure A2.
Ambari module and the results of the executed query

Appendix 3



Notes: Number of times that the recommended books have been borrowed from the University library in 2017. When the prototype of the Big Data ecosystem for the smart library has been enabled

Figure A3. Ambari module and the results of the executed query

Corresponding author
Aleksandar Simović can be contacted at: asimovich@gmail.com