ICIC Express Letters
Volume **13**, Number **10**, October **2019**
 ISSN 1881-803X
pp. 979–986

A FRAMEWORK FOR KNOWLEDGE MANAGEMENT SYSTEM WITH MAPREDUCE APPROACH TO OVERCOME INFORMATION OVERLOAD

Karto Iskandar[1,3], Harjanto Prabowo[2], Raymond Kosala[1] and Agung Trisetyarso[3]

[1]Computer Science Department, School of Computer Science
[2]Management Department, BINUS Business School Undergraduate Program
[3]Computer Science Department, BINUS Graduate Program – Doctor of Computer Science
Bina Nusantara University
Jl. K. H. Syahdan No. 9, Kemanggisan, Palmerah, Jakarta 11480, Indonesia
{ karto_i; harprabowo; RKosala; atrisetyarso }@binus.edu

Received February 2019; accepted May 2019

Abstract. *In the digital information era, knowledge develops significantly in the Knowledge Management System (KMS). In this corridor, the growth of digital content in the system seems to have a positive impact on the organization. However, this continued growth is not always beneficial for the organization over time. In this context, the condition is renowned as Information Overload (IO) in the system. The research question in the study is to determine the KMS framework to overcome the information overload issue. As this backdrop, the paper provides the KMS framework guidance to overcome IO issues based on the MapReduce approach. In this respect, this study aims to be a useful reference framework for organizations that address IO problems. The method in this study refers to literature studies which collect various references related to KMS and MapReduce algorithm. It directs to the use of in-depth observation to build the framework. Ultimately, this study concludes that the production of KMS MapReduce framework has been completed and set prepared for the implementation in a real case.*
Keywords: Knowledge Management System (KMS), Information Overload (IO), MapReduce approach, KMS MR framework

1. **Introduction.** Previous knowledge management science has long been introduced by Nonaka [1] and continued recently. Bearing this in mind, Knowledge Management System (KMS) science has also been long introduced by various researchers and simultaneously continued. Various studies related to KMS science take place in numerous countries and multiple areas of the field, such as government, offices, education, military, healthcare, public organizations, and arts [2]. It indicates that science attracts more researchers where the implementation is widely used in various studies.

With this in mind, the digital information era has changed many things related to information processing resulting in a growing number of studies about Information Overload (IO) [3-6]. IO science has long been coined by Jacoby et al. since 1974 through their papers [7]. The discussion related to this topic provides continuous opportunities. Importantly, the phenomena of research related to the subject of IO are evolving along with technological developments, socio-cultural and current digitalization conditions [8-10]. Researches related to this topic have found many solutions that have been applied to various areas of fields with multiple methods and approaches [11].

In the same line of thought, IO represents a condition where users receive too much information more than they can process principally [3,12]. In this spectrum, additional

DOI: 10.24507/icicel.13.10.979

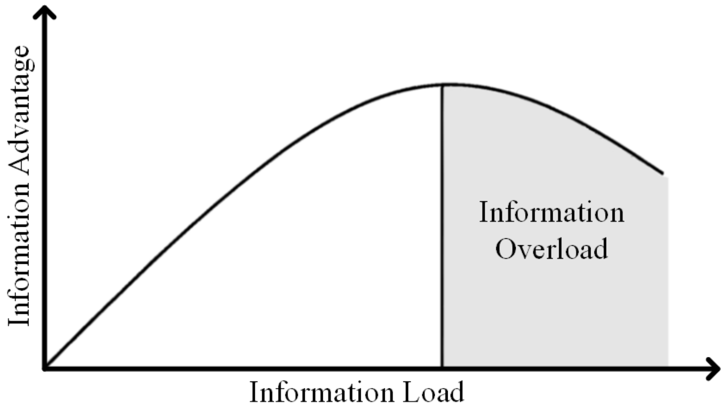

FIGURE 1. Information overload curve

information load contributes no impacts on information advantage for the users. Figure 1 describes IO graphically as an inverted-u curve [3,6].

Figure 1 depicts the illustration of IO in general where the addition of information load (horizontal axis) will have the positive impact on the information advantage (vertical axis) at the first moment to the peak. Hereafter, the additional information no longer has a positive impression or saturated, even will decrease [3,13]. In this case, the shaded area in Figure 1 describes the condition of information advantage that represents the information overload. On the vertical axis, the information advantage poses the benefits gained from information usages, such as decision accuracy [3], decision effectiveness [14], performance [15], and productivity [6].

With this over time, a system is continually operated actively so that a lot of data and documents are collected massively. It occurs in the implementation of KMS in an organization [16]. Digital content and organizing it in KMS is a significant concern and increase now and in the future. The world is currently in a state of information overload so that the information obtained on the system transcends the capacity that can be processed by humans [17]. Therefore, we need a method that enables to manage digital information in KMS and to overcome IO problems.

Various prevailing techniques and contemporary methods are used to maximize KMS in extracting large amounts of invaluable information and multiple structures to the organization. New ways, such as data mining processing, artificial intelligence, and extensive data processing, are still evolving. It is an exciting challenge, especially when an organization requires to maximize KMS by expecting the system able to handle information overload problems [18].

The research problem in this study addresses the type of KMS framework that assists organizations in overcoming the information overload on their KMS. On that basis, the framework is expected to be user-friendly and effortlessly implemented by organizations, so they can gain benefit from their KMS.

This study emphasizes on the significance of the KMS framework with different approval and assuredly solves IO problems. Therefore, the approach of this framework utilizes a MapReduce. The MapReduce algorithm was designed to overcome big data problems to overcome IO problems in KMS.

2. **Related Work.** In previous research, different studies underscored the process of tackling the information overload problem on a system as stated by researchers [19-21].

In this respect, Demirsoy's research explored how to overcome the information overload problems using the semantic knowledge management system approach. In the study, the researcher investigated the different methods of information retrieval models based on knowledge management systems in large-scale organizations from the perspective of software engineers. At the end of the study, the researcher concluded that the semantic knowledge management system signifies a very high potential to solve information overload problems in software engineering if the necessary measures are taken [19].

From another angle, Lee et al. discussed open innovation communities (MyStarbucksIdea.com) that face the challenge of information overload incurred due to the characteristic of the community. The research objective is to mitigate the IO problem in the open innovation environment. The researchers analyzed a large dataset collected from the communities utilizing text mining techniques including TF-IDF and sentiment analysis. It also considered both term and non-term features of the dataset. The results showed that term and non-term features play essential roles in predicting the hybrid classification models and achieved the adaptability of ideas and the best classification accuracy. In the end, the researchers found a recommendation system to mitigate the IO problem [20].

Per this outcome, Chen explained about how to reduce IO problem through improving website structure. It is important because a website evolves, the need for information also changes. This study uses the Mathematical Programming (MP) model to solve IO effectively eventually [21].

3. **Methodology.** The research method is a procedure for how research is conducted [22]. This study uses a combination of qualitative and observation techniques. As stated, the qualitative approach is used to find the answer to a research question by collecting various literature related to KMS, IO and MapReduce. Then in-depth observation of the collected documentaries intends to create a KMS framework using the MapReduce algorithm approach.

This study continues and extends the Study Literature Review (SLR) that has been done previously to find IO solutions. In the study, it proposes several solutions to IO problems, i.e., the use of information processing, information searching, application utilization, self/behavior-management, information filtering, information reduction, recommendation system, ontology, and classification [11]. Based on those results, the researchers created a KMS framework.

Not surprisingly, MapReduce becomes vital in research because it is a programming algorithm developed to expedite the massive data processing [23,24]. At first, MapReduce was released by Google's research team aiming at large-scale processing data distributed and parallel in clusters consisting of thousands of computers [25,26]. The application of the MapReduce algorithm in data processing is conducted by dividing the data according to the number of threads and entering data into the database and then displaying the results of the synchronization data [27,28]. In this point, it serves as the background for the selection of the MapReduce algorithm to build a KMS framework in this study. For this reason, the MapReduce algorithm is capable of processing an extensive database.

4. **Study Result.** The section presents the study result in the form of a KMS framework. It implemented a MapReduce algorithm approach. Based on several papers related to KMS, IO and MapReduce collected, an investigation was conducted to establish the framework. Furthermore, this framework is named KMS-MR framework. The KMS-MR framework is shown in Figure 2.

In this KMS framework, map and reduce processes occur after the subject experts (knowledge worker) perform the retrieval resources, which the knowledge source may come from external, internal, and personal resources. External resources referred to the sources from outside the organization, such as from Internet, books, magazines, articles, and journals. Equally important, internal sources represented the sources that already exist in the organization, such as from libraries, internal collection documents or materials, and KMS databases. In the end, personal resources are personal collections from experts, such as self-made documents or private collections.

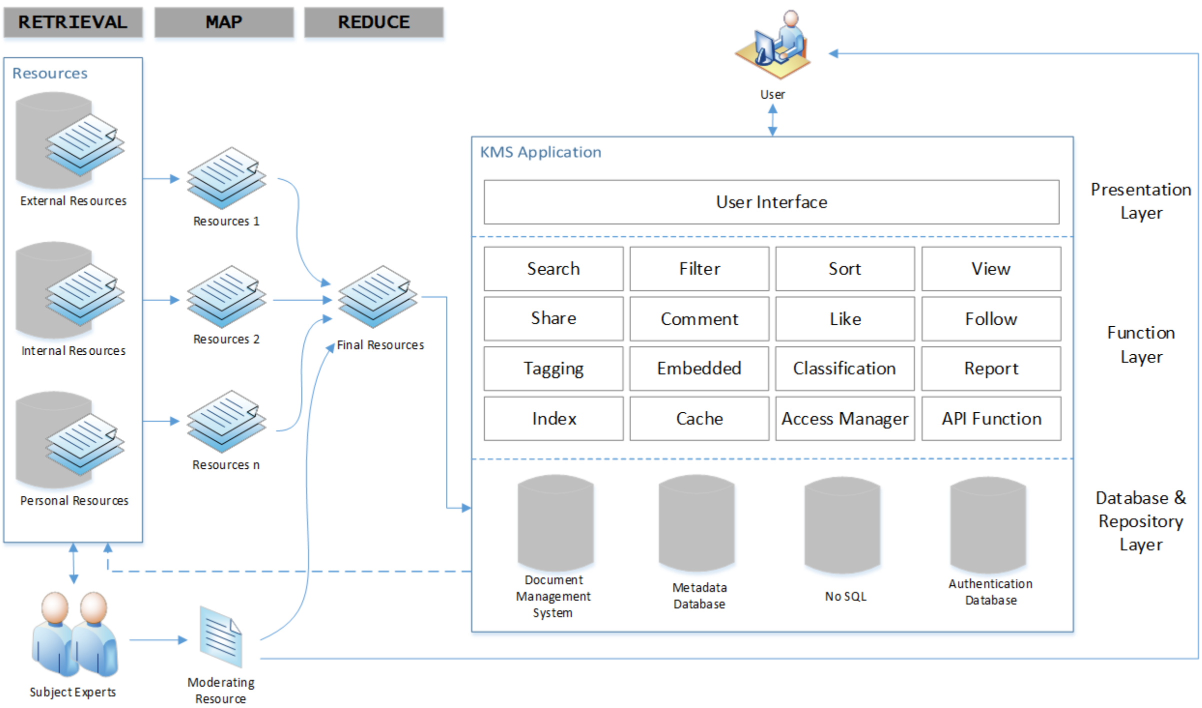

FIGURE 2. KMS MR framework

At the following procedure, all collected resources (primary resources) are entered to the MapReduce process. First, in the Map process, all resources are mapped based on the values or categories or other definitions required by the organization. The Map process will produce grouped or mapped resources, thus facilitating the next Reduce process.

Later on, that previous resources and added moderating resources became input to the Reduce process. Moderating resource is a set of resources to trigger the process. In practice, moderating resource includes the forms, such as accreditation documents, requirement projects, hypothesis document, exam documents, and a collection of problems documents.

Subsequently, in the Reduce process, grouped resources will be selected according to the needs criteria and moderating resources. At last, these results will be re-entered into the KMS database and could be used in the KMS application.

The KMS application model in this framework consists of three layers, i.e., the presentation, function, and the database & repository layer. Each layer has its role in this KMS application. The presentation layer is responsible for managing the display that interacts with the user directly. This layer contains the user interface component that is responsible for the user experience.

The next layer under the presentation layer is the function layer. This layer contains functions in the KMS application. There are sixteen functions, i.e., search function, filter function, sort function, view function, share function, comment function, like functions, follow function, tag function, embedded function, classification function, report function, index function, cache function, access rights function, and API (Application Programming Interface) function. The explanation for each function can be seen in Table 1.

The last layer in the KMS application is the database & repository layer. This layer has a function as a place to store all content, files, metadata and authentication database in the KMS application. In this layer, there are four databases, i.e., a document database, a metadata database, NoSQL database, and an authentication database. The explanation for each database in this framework can be seen in Table 2.

Table 1. Function layer components in the KMS application

No.	Function name	Function description
1	Search function	To retrieve information on a database or system based on input keywords. Generally, the search function focused on the retrieval and found documents most relevant to individual queries. Furthermore, there are two search methods, namely simple search and advanced search. Here simple search generally uses a text box for searching, while advanced search uses several fields for searching.
2	Filter function	To refine a set of the database on a range of data or based on desired criteria. For example, filter out a group of literature databases that are published from 2000 to 2018.
3	Sort function	To sort unsorted data to make it easier and faster for users to find or process the data. The most frequently used sorting are numerical and alphabetical sort. On the other hand, the sort of ascending and descending is also widely used based on user needs.
4	View function	To display data, to read documents, and to play multimedia files by users so that they can utilize the information. Occasionally this function is also equipped to view the metadata of the document.
5	Share function	To set a data or document to another user so that the user can joint to use and access the source. This share function has become popular and widely used especially on social media recently.
6	Comment function	To make comments on the data or document that are accessed by users. This comment is readable by other users so that fellow users can post a comment on each other. This comment function is useful for document owners to obtain feedback about the content.
7	Like function	To give a sign that users like the content or vice versa. This 'like' function has become popular in the social media era and is the most widely used recently.
8	Follow function	To follow a data, document or another user so that followers receive the notification. This following function has become popular and widely used especially on social media recently. A document that is followed by a lot of users indicates that it is more valuable.
9	Tagging function	To mark a data or document and to facilitate the use or access of the data or document in the future. Tagging function is also used to group the data or documents so that it is easy to process.
10	Embedded function	To include data or documents or multimedia file from the system to another system (embedded). This function enables other systems to use existing sources without copying the source. This function is perfect for avoiding duplicate resources.
11	Classification function	To group the data or documents that are presented so that it will make it easier for users to process them.
12	Report function	To present a report in the system such as the most frequently accessed document report, new document report, or another required report by user needs.
13	Index function	To run automatically in the backend. In general, this function is triggered by new content created or deleted from the system. This index function is beneficial for improving system performance, especially search features.
14	Cache function	To manage memory processing as a place to store temporary data needed by the processor. The cache functions to accelerate data access on the computer because the cache stores data or information that has been accessed by a buffer, thereby reducing the work of the processor.
15	Access management function	To manage user logins, user right to enter any module, and user right to access any data or documents. This function is essential in the KMS application to restrict documents or authorized information users.
16	API (Application Programming Interface) function	To facilitate programmer to exchange the information or data or documents so that these resources can be used by other systems.

Table 2. Database layer component in the KMS application

No.	Database	Database description
1	Document Database	A collection of digital documents or files that are stored systematically in a computer that can be managed or manipulated, and used by users by access authorization. Further, the collection can be in the form of office documents, read-only documents, video files, audio files, pictures, link documents, interactive documents and others. Each collection form has different characteristics and stores various information.
2	Metadata Database	A collection of information/metadata that describes other data (documents). Metadata database is defined as a database that stores the data providing information about one or more aspects of the data from the document database. An important metadata database enables users to rediscover, reuse, or manage documents efficiently. The metadata database generally consists of text forms and smaller sizes than the document itself. It reveals that index function is vital in this metadata database to speed up its performance.
3	NoSQL Database	A collection of information stored in the database with non-relational data types is useful for storing data that grows continuously. This NoSQL database is optional for modern KMS applications that contain semi structures and big data. The advantages of implementing NoSQL database in the KMS application impact on some operations faster.
4	Authentication Database	A collection of user, password, authentication and access rights to documents and systems. The authentication system can be either Active Directory (AD), Lightweight Directory Access Protocol (LDAP), or a simple database authentication system.

Based on the framework, it ushered that the development into pseudocode or algorithm aims to effortlessly implementation. Presumably, this pseudocode or algorithm is a sequence of logical steps to solve KMS problems with the MapReduce method. In this vein, it is arranged systematically and logically. This pseudocode is structured as a high-level description that is intended to readiness level by users so that it is implementable in a real case consequently.

5. **Conclusions.** The knowledge management system framework is still developing and adapting to the era and problems encountered. The principal challenge in the digital age of information today is the information overload phenomenon that occurs in various fields and various systems. Therefore, this KMS framework is developed by approaching the workings of the MapReduce algorithm and could overcome information overload by placing map and reducing processes carried out by subject experts before the sources entered in the KMS database.

It shed light on KMS framework that is equipped with a KMS application model consisting of three layers, namely the presentation layer, function layer, and the database & repository layer. These layers can be implemented on various distinct machines so that the performance of each layer can be monitored.

For further researches, it suggests adding additional methods, such as Artificial Intelligence (AI) to analyze the knowledge used and knowledge trace log so that the system is capable of providing recommendation knowledge to users according to their needs. Moreover, the framework achieved in this research recommends furthering research that implements the framework into real case studies. For the more comprehensive result, evaluations can be carried out on case studies from various field disciplines ultimately.

Acknowledgement. This research is supported by Doctor of Computer Science and School of Computer Science, Bina Nusantara University. Thanks to Pak Kristianus Oktriono from Language Center, Bina Nusantara University for proof-reading this paper.

REFERENCES

[1] I. Nonaka, A dynamic theory of organizational knowledge creation dynamic theory knowledge of organizational creation, *Organ. Sci.*, vol.5, no.1, pp.14-37, 1994.
[2] K. Iskandar, M. I. Jambak, R. Kosala and H. Prabowo, Current issue on knowledge management system for future research: A systematic literature review, *Procedia Comput. Sci.*, vol.116, pp.68-80, 2017.
[3] M. J. Eppler and J. Mengis, The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines, *IEEE Eng. Manag. Rev.*, vol.38, no.1, p.3, 2004.
[4] K. Mostak and G. Hoq, Information overload: Causes, consequences and remedies: A study, *Philos. Prog.*, vols.55-56, 2014.
[5] A. Barnea, Challenging the 'lone wolf' phenomenon in an era of information overload, *Int. J. Intell. CounterIntelligence*, vol.31, no.2, pp.217-234, 2018.
[6] P. Karr-Wisniewski and Y. Lu, When more is too much: Operationalizing technology overload and exploring its impact on knowledge worker productivity, *Comput. Human Behav.*, vol.26, no.5, pp.1061-1072, 2010.
[7] J. Jacoby, D. E. Speller and C. A. Kohn, Brand choice behavior as a function of information load, *J. Mark. Res.*, vol.11, no.1, pp.63-69, 1974.
[8] D. Allen and T. D. Wilson, Information overload: Context and causes, *New Rev. Inf. Behav. Res.*, vol.4, no.1, pp.31-44, 2010.
[9] M. I. Hwang and J. W. Lin, Information dimension, information overload and decision quality, *J. Inf. Sci.*, vol.25, no.3, pp.213-218, 1999.
[10] J. B. Schmitt, C. A. Debbelt and F. M. Schneider, Too much information? Predictors of information overload in the context of online news exposure, *Inf. Commun. Soc.*, vol.4462, pp.1-17, 2017.
[11] K. Iskandar, H. Prabowo, R. Kosala and A. Trisetyarso, The solution threshold of information overload: A systematic literature review, *ICIC Express Letters*, vol.12, no.12, pp.1223-1233, 2018.
[12] J. Jacoby, Perspectives on information overload, *J. Consum. Res.*, vol.10, no.4, pp.432-435, 1984.
[13] N. K. Malhotra, Reflections on the information overload paradigm in consumer decision making, *J. Consum. Res.*, vol.10, no.4, p.436, 1984.
[14] A. Intezari and S. Gressel, Information and reformation in KM systems: Big data and strategic decision-making, *J. Knowl. Manag.*, vol.21, no.1, pp.71-91, 2017.
[15] M. S. Owlia, A framework for quality dimensions of knowledge management systems, *Total Qual. Manag. Bus. Excell.*, vol.21, no.11, pp.1215-1228, 2010.
[16] J. L. Zhao, S. Fan and D. Hu, Business challenges and research directions of management analytics in the big data era, *J. Manag. Anal.*, vol.1, no.3, pp.169-174, 2014.
[17] M. A. Islam and M. Ikeda, Convergence issues of knowledge management in digital libraries: Steps towards state-of-the-art digital libraries, *Vine*, vol.44, p.7, 2014.
[18] L. Halawi, R. V. McCarthy and J. E. Aronson, An empirical investigation of knowledge management systems success, *J. Comput. Inf. Syst. Winter*, vol.48, no.2, pp.121-135, 2008.
[19] A. Demirsoy, *Using Semantic Knowledge Management Systems to Overcome Information Overload Problems in Software Engineering*, Master Thesis, Blekinge Inst. Technol. Sweden, 2013.
[20] H. Lee, K. Choi, D. Yoo, Y. Suh, S. Lee and G. He, Recommending valuable ideas in an open innovation community: A text mining approach to information overload problem, *Ind. Manag. Data Syst.*, vol.118, no.4, pp.683-699, 2018.
[21] M. Chen, Improving website structure through reducing information overload, *Decis. Support Syst.*, vol.110, pp.84-94, 2018.
[22] R. Kumar, *Research Methodology: A Step by Step Approach for Beginners*, SAGE Publications, Singapore, 2014.
[23] J. Dean and S. Ghemawat, MapReduce: Simplified data processing on large clusters, *Commun. ACM*, vol.51, no.1, p.107, 2008.
[24] J. Powell, L. Collins, A. Eberhardt, D. Izraelevitz, J. Roman, T. Dufresne, M. Scott, M. Blake and G. Grider, 'At scale' author name matching with Hadoop/MapReduce, *Libr. Hi Tech News*, vol.29, no.4, pp.6-12, 2012.
[25] J. Chen, Y. Chen, X. Du, C. Li, J. Lu, S. Zhao and X. Zhou, Big data challenge: A data management perspective, *Front. Comput. Sci.*, vol.7, no.2, pp.157-164, 2013.

[26] Z. Zhou, G. Liu and L. Su, A new approach to detect epistasis utilizing parallel implementation of ant colony optimization by MapReduce framework, *Int. J. Comput. Math.*, vol.93, no.3, pp.511-523, 2016.

[27] S. del Ro, V. Lpez, J. M. Bentez and F. Herrera, A MapReduce approach to address big data classification problems based on the fusion of linguistic fuzzy rules, *Int. J. Comput. Intell. Syst.*, vol.8, no.3, pp.422-437, 2015.

[28] H. Shi, X. Bai and J. Duan, 3D fabric dynamic simulation based on MapReduce, *Int. J. Cloth. Sci. Technol.*, vol.27, no.6, pp.793-802, 2015.