Initial Infrastructure for Knowledge Grid Model to Facilitate Knowledge Sharing in Big Data Environment

Sara Hosseinioun[1], Rusli Abdullah[2], Marzanah A. Jabar[2] and Yusmadi Yah Jusoh[2]

[1]*Universiti Putra Malaysia, Malaysia, {sarahosseinioun@gmail.com}*
[2]*University Putra Malaysia, Malaysia, {(rusli, marzanah, yusmadi)@upm.edu.my}*

ABSTRACT

Recently, the improvement of computer processing ability, storage capacity and, communication drive business and organizations to improve their resources by using big data, also they gathered by their interest as a community to share their knowledge and skills to develop vision and strategies. Knowledge grid designed to deal with a large number of data sets in grids for efficient capture, publish, share, and manage knowledge resources. This study discovered the relation between knowledge grid and big data and its impact on knowledge sharing. Also listed the factors which impact knowledge sharing from several models. It concluded with the connection of knowledge grid with these factors and community member's role in knowledge sharing among big data community.

Keywords: Knowledge Grid, Knowledge Sharing, Big Data Community.

I INTRODUCTION

In many scientific and business areas the huge pool of data needs to analyze and flow between the users. These analyses become more complicated while structured and unstructured data mixed with each other and their quantity increase in every second. This ocean of data which might store from all electronic devices around the world called big data. On the other hand, it has the ability to answer any questions and solve many problems in it which must extract by experts whom known as a data scientist. These group of big data users comes together by merging and supporting knowledge management system characteristics as big data community to help capture and share expertise, experiences, ideas. Thus, their communication and sharing knowledge which includes knowledge transferring and knowledge receiving (Hendriks P., 1999) is fundamental to community existence. Sharing data and information needs to be balanced and controlled to maximize its effect (Al-Nuaimi E., Al-Neyadi H. , Mohamed N. and Al-Jaroodi J., 2015) and moving huge amount of data need more workload of security to avoid any threat in transferring way and reach to the authorized member (Chen C.L.P. and Zhang C.Y., 2014). So developing an infrastructure to distribute geographically high performance discovered knowledge among the community is necessary. Knowledge grid offers high-level tools and techniques to distribute mined and extracted knowledge from data resources (Cesario E., Congiusta A., Talia D. and Trunfio P., 2007). It focuses on several type ways for knowledge might distribute as public knowledge share for several grids, sets of private knowledge send by a specific member to the group and finally a private knowledge share for a specific user (Zhuge H., 2002). It also able to intelligently clustered and fused knowledge by it relevancy and pervasive knowledge sharing while knowledge isn't statically stored and it can dynamically evolve to keep up-to-date.

This study focused on knowledge grid connection with big data in the first session. It also discovers the way it can improve and effect knowledge sharing among big data community from the extracted factors which impact on knowledge sharing.

II LITERATURE REVIEW

A. Knowledge Grid

Knowledge Grid defined as communication infrastructure (Wang G., Wen T., Guo Q. and Ma X., 2006), which use for mining large set of data available over the grid and improve organizational models and industrial processes (Cannataro M. and Talia D., 2004).Moreover, It has three characteristics as networkability, inter-operability and composition ability (Zhuge H., 2002), that makes resources around the world as grid nodes connected semantically by grid and enables users to make references, answer question and find conclusion from masses of data (Wang G., Wen T., Guo Q. and Ma X., 2006). It designed to implement data mining applications and services which includes resource management that provides a system to describe, publish information from data sources, computing resources and data mining algorithms (Cesario E., Congiusta A., Talia D. and Trunfio P., 2007).

B. Big Data Community

Big data defined as a large pool of data that is beyond the current ability to process (Chandhini C. and Megana L.P, 2013). Individuals and

organizations from different areas health and medical, commerce and business, science and education, social media and etc. around the world use big data to enhance the quality and quantity of their services and product and improve their decision making. So these big data users linked like members of the community as a small or large unit which it members have something in common; such as virtual community approach that is an effective term for sharing knowledge available from user experiences and skills with other members (Lin H. F., 2008). However, from the huge variety of big data community members, the research focused on data scientist whom should handle all the big data challenges and able to bring structure to large quantities of formless data and make analysis possible (Davenport T.H. and Patil D.J. , 2012). So they must be able to discover knowledge that meets the defined need or provide the answer to the defined problem. In order to operationalize this knowledge, a template can help them to obtain the benefits of information discovery and share it while avoiding the rediscovery (Englmeier K. and Murtagh F. , 2017).

C. Knowledge Grid among Big Data

Knowledge grid represents and exchanges effectively massive data and information while integrates and converts them into useful knowledge through mining and reference methods (Wang G., Wen T., Guo Q. and Ma X., 2006). In addition, big data main characteristics as scale or volume, variety and speed or velocity are mixed in knowledge grid structures. It initially designed for dealing with problems involving large amounts of data or compute-intensive applications (Aminu, L. M., 2014).While knowledge resource as grid nodes connected through the grid is able to mine a huge quantity of the different type of data sets in grids (Cannataro M. and Talia D., 2004). A grid is an effort to create an advanced infrastructure, targeting at an adaptive vast resource environment, integrating higher-level services which enable applications to adapt to heterogeneous and dynamically changing meta-computing environments, with ease, low cost, reliability, and regardless of location and devices (Zhuge H. and Sun X., 2016). In addition, the knowledge resources as grid nodes and the semantic connection between them include grid network as grid links are main components of knowledge grid which has direct interaction with users whom are big data generators and define the semantic relation between knowledge.

III   KNOWLEDGE GRID AND KNOWLEDGE SHARING AMONG BIG DATA

To analyze the connection between knowledge grid and knowledge sharing through big data community this section first analyzes the factors from several research models for facilitating knowledge sharing. In addition, studied knowledge grid components and their impact on knowledge sharing.

A. Factor Influence Knowledge Sharing

Facilitating knowledge sharing is a big or general word and need to broke into several factors to understand how can enhance it. Also, it is essential to discover the factors effect knowledge sharing positively or negatively. Table 1 listed these factors from different research which they all conclude with research model for knowledge sharing.

Table 1. Factors influence Knowledge Sharing.

Author	Factor	Focus	Result
(Chena X., Li X.,Clark J. G. and Dietrich G.B., 2013)	Knowledge Differentiation	open source software project teams	Facilitate Knowledge Sharing and Communication Quality
	Knowledge Location		
	Knowledge Credibility		
(Siemsen E., Roth A. V. and Balasubramanian S., 2008)	Motivation	Organization employees	Improve knowledge sharing behaviors
	Ability		
	Opportunity		
(Chen C. J. and Hung S.W. , 2010)	Norm of Reciprocity	Professional virtual communities	knowledge contributing and collecting behaviors
	Interpersonal Trust		
	Knowledge Sharing Self-Efficacy		
	Perceived Relative Advantage		
	Perceived Compatibility		
(Alali H. and Salim J., 2013)	Knowledge quality	Healthcare sector in virtual communities	Improve knowledge sharing behavior
	System Quality		
	Service Quality		
	Perceived Usefulness		
	Perceived Ease of Use		
(Zheng Y., Zhao K. and Stylianou A., 2013)	Perceived Information quality	Virtual Communities	continuance intention to consume and continuance intention to provide
	Perceived system quality		
(Chiu C. M., Hsu M. H. and Wang E. T.G, 2006)	Social Interaction Ties	Virtual Communities	Quantity of Knowledge Sharing and Knowledge
	Trust		

	Norm of Reciprocity		Quality
	Identification		
	Shared Language		
	Shared Vision		

To analyze and categorize factors, tried to choose environment which is similar to big data community and the factor proofed by empirical analysis.

B. **Factor Influence Knowledge Sharing**

A grid in most of the researches explicated in the grid computing which is the technology makes large-scale distributed computing system able to share computing resources (Zhuge H., 2002). Grid computing refers to the many of servers as grid nods that interconnected in a high-speed network and play one or several tasks (Aminu, L. M., 2014). Moreover, grid network provides high-performance communication service with great fault tolerant ability (Chandhini C. and Megana L.P, 2013). Knowledge grid is on top of grid computing and grid network which analyze and connect information semantically (Zhuge H., 2004). Knowledge grid support data management and data analysis applications by offering resources and decentralized data access system to integrated services to backing distributed knowledge discovery in the grid (Cesario E., Congiusta A., Talia D. and Trunfio P., 2007).

IV   DISCUSSION

This study has reviewed the relationship between knowledge grid and big data community to improve knowledge sharing among big data community so it listed several factors which influence knowledge sharing that can categorize from several aspects.

Users are an important part of the big data system and their behavior influences big data generating and using (Zhuge H. and Sun X., 2016). Users or member's behavior and expectation aspect cover many of knowledge sharing factors. The behavioral view refers to the characteristic which gives the pleasure in helping people (Lin T.C., Wub S. and Lu C.T., 2012) and the intention and reliability level of members knowledge sharing (Chen C. J. and Hung S.W. , 2010). In the journey of digging the valuable knowledge from big data chaos, data scientists need to rely on each other experience and skills to avoid rediscovery so trust and reliability can drive members to share their knowledge. More interest in specific aspect has a direct impact on increasing knowledge generation which effects on the density of resources and grid nodes. Also should consider

that users understanding of the information create and define the semantic connection of knowledge influence semantic links in the grids. In addition, what community members expect from the system defined, user satisfaction or expectation as the dependent variable in many knowledge sharing models lead other groups of factors which can organize by their impact on quality of shared knowledge and distributing system.

From the knowledge quality view, shared knowledge should specialize in different knowledge domains. By considering big data characteristics, it should be accurate, believable, credible, understandable and sufficiently up to date to fulfill a specific need and provide advantage from its use (Zheng Y., Zhao K. and Stylianou A., 2013). Semantic analysis has the main role in understanding, organizing, computing and finally interpret of big data (Zhuge H. and Sun X., 2016).Therefore knowledge grid as a set of well organize knowledge and knowledge management operations with the ability to analyze semantically through the data sources can guarantee the accuracy of identified knowledge. Moreover, it can integrate grid sources to support all the knowledge discovery process by identifying the relationship between raw data content.

Ease of use of distributing system can affect the member's quantity of using system, its accessibility, security, flexibility and response time increase user satisfaction (Alali H. and Salim J., 2013). When a huge amount of data from data resources share and flow across different nodes, it should reach only to the authorized community members. Also, it's important to protect knowledge from modifying on its path. The Grid Security Infrastructure certificates for authentication and control who is allowed to access data, and protect it during distribution which can adopt different methods such as replication, encryption, source routing and etc. (Esposito V., Ficco M. and Palmieri F., 2016) Also knowledge grid able community members to access and manage knowledge and distribute it around the world through a single semantic entry point without knowing the required knowledge's locations. Thus, knowledge grid has potential to support most of the knowledge sharing factors effectively, however, it is essential to have an accurate understanding of the semantic relation and link strength between grid node as represent of knowledge resources by considering their diversity and density in distributing network to establish suitable knowledge grid model for efficient knowledge sharing.

V   CONCLUSION

Currently, the effort of using the BD resources to improve decision making, services or products increase significantly. So data scientists whom able to extract valuable knowledge from big data can categorize as a group of big data community member which can share their experiences and skill among the community. While sharing knowledge through the community because of big data characteristics are complicated, knowledge grid as distributed infrastructure can connect data resources semantically. This study tried to discover the initial fundamental for designing a knowledge grid model which can facilitate knowledge sharing. To achieve the objective, the factor can influence knowledge sharing to categorize by their result from several models which defined in the environment similar to big data community. Finally, the research connected the knowledge grid ability and characteristics to the knowledge sharing factors which analyzed from community member aspect and their satisfaction from the knowledge sharing system, how knowledge grid can affect the quality of knowledge and the expected quality of a distributing system.

REFERENCES

Alali H. and Salim J. (2013). Virtual Communities of Practice Success Model to Support Knowledge Sharing Behaviour in Healthcare Sector. *Procedia Technology , 11*, 76 – 183.

Al-Nuaimi E., Al-Neyadi H. , Mohamed N. and Al-Jaroodi J. (2015). Applications of big data. *Journal of Internet Services and Applications, 6*(1), 1-15.

Aminu, L. M. (2014). Implementing Big Data Management on Grid Computing Environment. *International Journal Of Engineering And Computer Science, 3*(9), 8455-8459.

Cannataro M. and Talia D. (2004). Semantics and knowledge grid: Building the next generation grid. *IEEE Computer Society, 4*, 56-63.

Cesario E., Congiusta A., Talia D. and Trunfio P. (2007). Designing data analyis services in the knowledge grid. *Core Grid*, 1-15.

Chandhini C. and Megana L.P. (2013). Grid computing a next level challenge with big data. *Interenational journal of scientific & engineering research , 5*(3), 1-5.

Chatfielda A. T. and Reddick C. G. (2017). Customer agility and responsiveness through big data analytics for public value creation: A case study of Houston 311 on-demand services. *Government Information Quarterly*, 1-12.

Chen C. J. and Hung S.W. . (2010). To give or to receive? Factors influencing members' knowledge sharing and community promotion in professional virtual communities. *Information & Management, 47*, 226–236.

Chen C.L.P. and Zhang C.Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences, 275*(1), 314–347.

Chena X., Li X.,Clark J. G. and Dietrich G.B. (2013). Knowledge sharing in open source software project teams: A transactive memory system perspective. *International Journal of Information Management, 33*, 553–563.

Chiu C. M., Hsu M. H. and Wang E. T.G. (2006). Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories. *Decision Support Systems, 42*, 1872–1888.

Davenport T.H. and Patil D.J. . (2012). Data scientist: The sexiest job of the 21st century. *Harvard Business Review, 90*(10), 70-76.

Englmeier K. and Murtagh F. . (2017). What Can We Expect from Data Scientists? *Journal of theoretical and applied electronic commerce research, 12*(1).

Esposito V., Ficco M. and Palmieri F. (2016). On Reliability and Security in Knowledge Grids., (pp. 1-8).

Hendriks P. (1999). Why share knowledge? The influence of ICT on the motivation for knowledge sharing. *Knowledge and Process Management, 6*(2), 91-100.

Li X., Wu P., Shen G. Q., Wang X. and Teng Y. (2017). Mapping the knowledge domains of Building Information Modeling (BIM): A bibliometric approach. *Automation in Construction, 84*, 195–206.

Lin H. F. (2008). Determinants of Successful Virtual Cmmunities: Contribution from system Characteristics and Social Factors. *Information & Managment, 45*, 522-527.

Lin T.C., Wub S. and Lu C.T. (2012). Exploring the affect factors of knowledge sharing behavior: The relations model theory perspective. *Expert Systems with Applications , 39*, 751–764.

Siemsen E., Roth A. V. and Balasubramanian S. (2008). How motivation, opportunity, and ability drive knowledge sharing: The constraining-factor model. *Journal of Operations Management , 26*, 426–445.

U. Sivarajah, M. M. Kamal, Z. Irani, V. Weerakkody. (2017). Critical analysis of Big Data Challenges and Analytical Methods. *Journal of business research, 70*, 263-286.

Wamba S. F., Gunasekaran A., Akter S., Ren S. J., Dubey R. and Childe S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research, 70*, 356–365.

Wang G., Wen T., Guo Q. and Ma X. (2006). A knowledge grid architecture based on mobile agent. *IEEE computer society.*

Zheng Y., Zhao K. and Stylianou A. (2013). The impacts of information quality and system quality on users' continuance intention in information-exchange virtual communities: An empirical investigation. *Decision Support Systems, 56*, 513–524.

Zhuge H. (2002). A Knowledge grid model and platform for global knowledge sharing. *Expert systems with Applications, 22*, 313-320.

Zhuge H. (2004). China's E-Science Knowledge Grid Environment. *IEEE computer Society, 4*, 13-17.

Zhuge H. and Sun X. (2016). Semantics, Knowledge and Grids on Big Data. *Future Generation Computer Systems, 64*, 163-164.