International Journal of Engineering & Technology, 7 (4.5) (2018) 631-633

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Big data and semantic web, challenges and opportunities a survey

Jeelani Ahmed [1] *, Dr. Muqeem Ahmed [2]

[1] *Research Scholar Dept. of CS&IT, Maulana Azad National Urdu University, Hyderabad*
[2] *Assistant Professor, Dept. of CS&IT, Maulana Azad National Urdu University, Hyderabad*
**Corresponding author E-mail: muqeem.ahmed@gmail.com*

Abstract

In recent years, vast and complex amounts of data are being created and making it difficult for traditional data processing applications to manage them. The coming of the Internet prompted monstrous spike in the volume of information being made and made accessible. World Wide Web consortium W3C and international standardization body of the web spread the Semantic Web. It is an extended form of current web which provide easier way to search, reuse, combine and share information. In the last few years, major businesses corporations have demonstrated interest in incorporating semantic web technology with big data for added value. Indeed this incorporation has some benefits as well; it increases end-users ability to self-manage data from various sources, it on focuses changing business environments and varying user needs and handles concepts and relationships, manages terminology while connecting different data from varied data sources. For Social Network Analysis (SNA) new methods are needed by combining Big Data and Semantic Web technologies as a way to utilize and add capacities to existing frameworks. Moreover, the fast changing business requirements and latest industry culture of Agile Development needs a robust yet flexible solution for Business Intelligence and by using distributed enterprise level ontologies Data Warehousing can be incorporated. This paper is an attempt to focus on effects of incorporating Big Data with Semantic web, how Semantic Web making Big Data smarter, revisit the Big Data and Semantic Web challenges and opportunities, relationship between them and finally we summarizes with future direction of this integration

Keywords: *Big Data; Semantics Web; Ontologies.*

1. Introduction

The recent growth of business processes and electronic data management brought a number of improvements for enterprises, such as selling products, automatic handling of purchasing. Therefore, information about business processes and products are managed as data in enterprises information systems [11]. However, by the increase of complexity required to handle more and more business processes and electronic data; enterprises are challenged by these complexities. Frequently, this issue is discussed in context of big data. Big Data has been getting increasingly attention and recognition due to its vast studies and application prospects [1], [2], [4]. Large-scale data and rapidly changes in many types are generally referred as Big Data. In Big Data, the data sets are usually collected and integrated from distinctive sources that may comprise of structured, semi-structured, and unstructured data [9-10]. And most of the data is unstructured and semi-structured data usually more than 85% [7]. There is semantic heterogeneity and structural heterogeneity problems involving in the data sources because of Big Data applications often require multiple data sources. Structural heterogeneity of data refers to different data stored in the data model that could not be directly mapped to each other. Data that is inconsistent with each other, unable to reflect the link between the data sets and could not understand each other is referred as Semantic heterogeneous data [11].

However, as product complexity increases and customers demand for new customizable products grows then companies have to handle more and more complex information about the products. Also, in order to handle and make use of large and complex data (big data) many companies faces the challenges of integrating vast amount of data that is disordered across many distinct information systems into one enterprise-wide centralized information system [6].

Today, many organizations are not ready to implement such an information framework into their information technology structure as procedures for information and business process integration are still too exorbitant and

Tedious for them. This can results in higher number of inconsistencies and data redundancy within the product information of an enterprise [5]. Hence, more expenses can arise for presenting, searching for and maintaining product information. Also, the number of wrong orders, wrong deliveries and customer requests may increase. Therefore, to avoid these wrong interpretations or redundant statements in complex and big data sets it is very important to implement syntactical, as well as semantically restrictions.

In this paper we present a brief survey of the technological aspects related to big data in terms of structured, semi structured and unstructured data from distinct sources and their inconsistencies with related to structural and semantic heterogeneity. In section II we described the architectural solutions for big data, paying particular attention to business processes and electronic data management. Section III discuss about semantic web technologies. The technologies that supports big data integration are discuss in section IV. And finally section V concludes with open problems and future challenges about this integration.

2. Big data management

Big Data is a term that depicts conceivably inconsistent data that lives in extensive volumes, diverse structures and is being generated at fast. According to such description, to overcome above such mentioned difficulties tools should be used that manage and operate on big data in a way to capture process and analyze the data. Big Data Management provides such devices and strategies to beat these challenges. Data Science Series (2012) gives a broadened rundown of conceivable advantages for the business and as well as for customers to turn into the big data resources. As the list depicts that any company independent of the sector can be get benefited from big data or specialty it involves as new open doors in information usage can be found and exploited As it had been stated, Big Data Management should use proper tools and strategies to make it conceivable to catch, process, and analyze down data that is fast, huge, and heterogeneous.

The present answer for Big Data Management that perhaps originates from appropriated sources is NoSQL databases. NoSQL databases are to a greater degree a philosophy instead of a technique or a tool. It portrays an arrangement of methods the Big Data Management can be expert. Data Storage in NoSQL database is completely different as compared to Traditional Relational Databases as shown in figure 2.1(a). For example, some NoSQL databases might utilize relation, a few try not to utilize SQL management language, and some may utilize schema- less, schema-free or flexible schema methods. Also, unique ways to deal with store information are being utilized. For instance, a few variety is key document framework some frameworks utilize key-value storage system, some uses graph families type or even column families type as shown in figure 2.1(b).

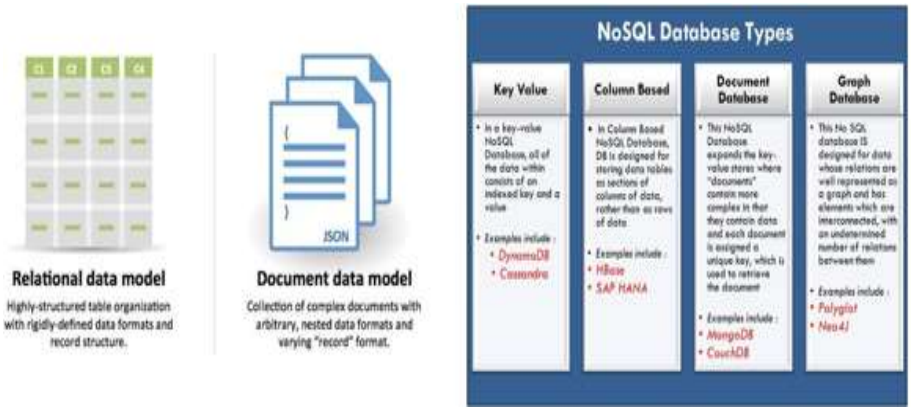

Fig. 2.1: A) Relational vs Document Data Model, B) NOSQL Database Types.

As discussed in a paper by DeCandia, Hastorun, Jampani, Kakulapati, Lakshman, Pilchin, Sivasubramanian, Vosshall &Vogels (2007) NoSQL utilizes a distributed architecture, with the information held in a repetitive way on a few hubs. By adding more servers the system can easily scale out and a server failure can be endured. This kind of database is utilized for huge amount of data management and these databases scales horizontally. By creating redundant copies of data, producing them strong to partition failure, high availability aspects of NoSQL databases are enhanced. ACID aspects can also be taken into consideration by NOSQL Databases. Burden places on databases may results that NOSQL does not provide complete consistency across distributed nodes. Many NoSQL databases can also be treated as schema-free databases. The key benefit of this design is that it allows various applications to upgrade the structure of data without rewriting the table. It also can provide greater flexibility to store heterogeneously structured data.

3. Semantic web technologies

Semantic web is a thought of adding significance to the things that are found on the World Wide Web. The motivation behind the additional importance is to enable machines to give purpose about these things. Ontologies on the other hand are the major driver behind the semantic web and define the way to integrate the data between heterogeneous information sources and can also be used to describe the connections between them. It is a standard that describe the naming, definitions, relations and properties of different entities in a particular domain. The semantic web advances the standard for the explanation and integration of data. By empowering the consideration of semantic content in data open through the Internet, the point is to change over the present web into a web of data, which is overwhelmed by semi-structured and unstructured documents. It includes publishing information in languages specifically designed for data: Resource Description Framework (RDF), SPARQL (which is a protocol and query language for semantic web data sources), Web Ontology Language (OWL), and Extensible Markup Language (XML) [11].

Oktie Hassanzadeh, Anastasios Kementsietsidis, Yannis Velegrakis (2012) described in a paper that the goal of semantic web is to intensify the present web by linking the data and enriching it with meta data in ways that facilitate both the exploitation semantics of data and understanding of data. This enables the web with new qualitative levels of service. Selection of the Resource Description Framework (RDF) [5] as the data model for exchange and representation of the new web of data was one of the key choice for the Semantic Web community, towards tending to its difficulties and accomplishing its objectives. In spite of the reputation of the relational, and the XML model and in spite of the maturity of these models as far as both research and commercial system support RDF was chosen.

4. Integration of big data using semantic web technologies

Data integration includes two principle tasks. The first, at the schema level, includes homogenizing contrasts in the diagrams and classification used to speak to the information. The second, incorporation at the information level, includes recognizing records in various informational collections that allude to a similar certifiable substance. Li Kang, Li Yi, LIU Dong (2014) analyzes the problem in the big data integration process interns of structural heterogeneity and semantic heterogeneity. To solve some research problems in integration process they proposed a big data semantic model based on ontology by combining the semantic web technology, by building the ontology between the semantic models, to solve the problem of data unintelligible. Based on ontology to solve the problem of heterogeneous data storage they constructed a Key/Value storage model. A new system called Karma has been proposed by [20] in a case study, that trying to solve the variety challenge by using the semantic technology to integrate different type of Big Data sources.

Starting from the importing different types of information sources, cleaning process, modeling and its problems and concluding with integration process , a detailed information of different steps have been given. The proposed system uses semantic RDF technology to resolve integration issues. For instance, identifying same entities in different datasets at schema Level. The system is validating only particular types of structural and semi-structural data sources.
According to [21], the emotional and sentimental analysis of social networks in the financial domain using a combination of social networks (Twitter), semantic financial ontologies (FIBO), and other assets that provides uniform vocabulary to express emotions and sentiments in a proper format. Different data like emotions, opinions, and activities gathered from Twitter with FIBO and other data assets provide better understanding of different communities, which may be an important thing in financial domain. The work is executed in a legitimate and sorted out way and the proposed framework is clarified in detail. There is an absence of assessment what's more, approval of the proposed strategies, and the conclusion is general and does not obviously mirror the advantages of the proposed approach. In field of medical science consider a patient, who need personalized health care service design [12]. To achieve this necessity, a medicinal services framework needs to actualize a new foundation that would permit live delivery of patient information straightforwardly under the control of an expert. The opposite side of the condition would permit health care services frameworks to improve decisions about their patients in view of the information from every one of the patients. The paper presented & examines an approach towards Personalized Medicine with Big Data and Semantic Web Technologies. Large amount of homogeneous and heterogeneous information about each single patient can be accessed and processed by Big Data as part of smart data. Since the information isn't organized more often than not, Semantic Web Technologies become possibly the most important factor and are utilized to explain different ideas.

5. Conclusion

The above-discussed cases explain that the recent research on semantic ontologies to integrate big data sources and incur with structured, semi-structured and unstructured data. Various data sources and approaches are used trying to get better results. And more of the works used limited and basic semantic ontologies and did not fully counter with the Big Data. Moreover, integration of Big Data using ontologies and semantic web technologies to solve the variety problems is at the early stages and still a more research challenge that need future work especially in the health care domain, business and financial domain, which will creates an opportunity to have a greater contribution to the science. Both the Big Data and Semantic Web Technologies can be viewed as two unique technologies of research, both are connected to genuine certain issues. Furthermore, it can be seen that the principle focus is knowledge for applications of both the technologies may it be for more benefit or for the revelation of more learning. Accordingly, it can be state that both areas should advance further by offering importance to huge, fast, unstructured and uncertain data around us.

Acknowledgments

This survey work is a part of my research in Big Data and I would like to pay special thanks to my supervisor.

References

[1] K. Davis, D. Patterson, Ethics of Big Data: Balancing Risk and Innovation, O'Reilly Media, 2012.
[2] Li Kang, Li Yi, LIU Dong, "Research on Construction Methods of Big Data Semantic Model", World Congress on Engineering 2014 Vol I, WCE 2014.
[3] Qudamah Quboa and Nikolay Mehandjiev, "Creating Intelligent Business Systems by Utilizing Big Data and Semantics", IEEE 19th Conference on Business Informatics, Volume 2, 39-46, 2017.
[4] G. Halevi, H. Moed, The evolution of big data as a research and scientific topic: Overview of the literature, Res. Trends (2012) 3–6.
[5] Boris Mocialov, "Big Data Management Assessed Coursework Two Big Data vs Semantic Web", Heriot-Watt University, Edinburgh, April 5, 2015.
[6] David Ostrowski, Nestor Rychtyckyj, Perry MacNeille and Mira Kim,"Integration of Big Data Using Semantic Web Technologies" IEEE Tenth International Conference on Semantic Computing, 382-385, 2016.
[7] Srividya K Bansal (2014), "Towards a Semantic Extract-Transform-Load (ETL) framework for Big Data Integration", IEEE International Congress on Big Data, 2014.
[8] Olivier Cur´e, Fadhela Kerdjoudj, Chan Le Duc and Myriam Lamolle, "On the Potential Integration of an Ontology-Based Data Access Approach in NoSQL Stores", Third International Conference on Emerging Intelligent Data and Web Technologies, 166-173, 2012.
[9] Srividya K Bansal and Sebastian Kagemann, "Integrating Big Data: A Semantic Extract-Transform-Load Framework" IEEE Computer Society,Volume 48, Issue 3, 42-50, 2015.
[10] Li Ma, Jing Mei, Yue Pan Krishna Kulkarni Achille Fokoue and Anand Ranganathan, "Semantic Web Technologies and Data Management", IBM China Research Laboratory, 2007.
[11] Ivan Merelli, Horacio Pérez-Sánchez, Sandra Gesing and Daniele D'Agostino, "Managing, Analysing, and Integrating Big Data in Medical Bioinformatics: Open Problems and Future Perspectives", BioMed Research International, Volume 2014, Article ID 134023, 13 pages, 2014.
[12] Maryam Panahiazar, Vahid Taslimitehrani, Ashutosh Jadhav, Jyotishman Pathak, "Empowering Personalized Medicine with Big Data and Semantic Web Technology: Promises, Challenges, and Use Cases", IEEE International Conference on Big Data, [Online] 978-1-4799-5666-1, 2014.
[13] Bastian Eine, Matthias Jurisch, and Werner Quint (2017), "Ontology-Based Big Data Management", System, 5, 45, 1-14, 2017
[14] Jing Xiong, Yuntong Liu and Wei Liu (2014), "Ontology-based Integration and Sharing of Big Data Educational Resources", 11th Web Information System and Application Conference, 245-248, 2014.
[15] Loukia Karanikola, Isambo Karali and Sally McClean (2014), "Uncertainty reasoning for the Big Data Semantic Web", IEEE 15th International Conference on Information Reuse and Integration, 147 – 154, 2014.
[16] Knoblock, Craig & Szekely, Pedro (2015), Exploiting Semantics for Big Data Integration. AI Magazine. 36. 25-38.
[17] Gary E. Marshall, Paul A Tibbits (2016), "Data Integration with Semantic Web Technologies (SWT)", office of technology strategies (TS), office of information and technology (OI&T), Version 1.0, 2016.
[18] C.Kacfah Emani, et al., Understandable Big Data: A Survey, Computer Science Review (2015), http://dx.doi.org/10.1016/j.cosrev.2015.05.002.
[19] Zhang, J., & Huang, M. L. (2013, December). 5Ws model for big data analysis and visualization. In Computational Science and Engineering (CSE), 2013 IEEE 16th International Conference on (pp. 1021-1028). IEEE.
[20] C.A. Knoblock, and P. Szekely, "Exploiting semantics for Big Data integration," AI Magazine, vol. 36(1), pp. 25-39, 2015.
[21] J.F. Sánchez-Rada, M. Torres, C.A. Iglesias, R. Maestre, and E. Peinado, "A Linked Data approach to sentiment and emotion analysis of twitter in the financial domain," Proceedings of the Second International Workshop on Finance and Economics on the Semantic Web (FEOSW 2014), Anissaras, Crete, Greece. 26th May 2014, pp.51-62, 2014.