# Creating Intelligent Business Systems by Utilising Big Data and Semantics

**Document Version**
Accepted author manuscript

**Citation for published version (APA):**
Quboa, Q., & Mehandjiev, N. (2017). Creating Intelligent Business Systems by Utilising Big Data and Semantics. In *IEEE 19th Conference on Business Informatics (CBI)* (Vol. 2, pp. 39-46) https://doi.org/10.1109/CBI.2017.71

**Published in:**
IEEE 19th Conference on Business Informatics (CBI)

# Creating Intelligent Business Systems by Utilising Big Data and Semantics

Qudamah Quboa

Alliance Manchester Business School
University of Manchester
Greater Manchester, UK
qudamah.quboa@manchester.ac.uk

Nikolay Mehandjiev

Alliance Manchester Business School
University of Manchester
Greater Manchester, UK
n.mehandjiev@manchester.ac.uk

*Abstract*—**Increasing volumes of data are generated from a variety of industries, organizations and people. Such data is now seen as a key business asset, since analysing it allows business decision makers to make choices with confidence. This data is often "big data" – a data of such volume, velocity and variety that conventional processing techniques are not suitable. We research the specific challenge of variety, where data comprises a number of incompatible formats. Semantic technology, or the use of ontologies, is seen as a core approach to resolving this challenge and aligning the data generated from heterogeneous data sources. This paper lays down the issues which should be addressed and reviews the work which has been done in the area of integrating structured and unstructured data sources, with a special focus on the financial domain. We then chart a research plan in the area following Design Science research.**

*Keywords-Big Data; data integration; Big Data analytics; semantic web; Ontology; business intelligence system*

## I. OVERVIEW

The world is increasingly being digitized and increasing amounts of data are generated from heterogeneous sources, reflecting the actions of industries, organizations and people for example, conducting online transactions [1], selling products, rating services [2] and collating users' opinions through social network platforms [3]. Understanding and utilizing data is a key to taking actions that could minimize business losses, detect market trends, increase product sales and give more options to industry decision makers to make choices with confidence [4].

When this data is of size and complexity which makes it unsuitable for processing with conventional analysis tools, we call this "Big Data". Examples of systems and tools which work with big data include intelligent transportation media retrieval systems [5], Sociopedia, social media event and trend detection [6], Stock market predication, a social media sentiment system [4], Statoil OBDA system, data exploration system for Statoil Petroleum Company [7] and Karma, data integration system [8].

Big Data is characterised by the three main challenges (the 3Vs): volume, velocity and variety [3]. Currently, Business corporations are considering the work on the variety challenge as the most important factor to success (69% of all respondents) compared to volume (25%) and velocity (6%) [9] and the variety is representing the major driver of business companies' Big Data investments (40%)

compared to volume (14.5%) and velocity (3.6%) via integrating more sources and types of data to create greater business insight as reported in the executive summary of Fortune 1000 [10]. Our research focuses on the variety challenge whilst considering the constraints of the other two challenges. Semantic technologies and ontologies are seen as a core approach to this.

Ontology is defined as "a formal representation of the concepts in some domain" [11]. As ontologies and semantic applications have been successfully used to construct common understanding and domain knowledge representation, they are considered a likely solution to deal with the variety challenge of Big Data and to align different data sources [12,8,13,14]. Our research aims to address the issues of variety of big data and to develop an approach that provides a better integration between structured and semi-structured data sources, ensuring integrated data and enabling the construction of enhanced business intelligence tools.

Achieving this aim will bring about the following artefacts: (1) an approach using semantics and ontologies to integrate Big Data; (2) integration algorithms underpinning the approach; and (3) a prototype tool to test the feasibility of (1) and (2).

The application of semantic technologies in the domain of Big Data integration is still new and there is a need to deal with these issues: (1) the representation of complex data (structure, volume, and data heterogeneity) [15]; (2) data access, search, and management [16]; (3) data processing dealing with dynamic changes in the data itself and its features [17]; (4) data analysis [18]; and (5) data testing and validation [19].

Our research will focus on the first issue (complex data representation), aiming at creating an integration approach that allows the development of more advanced business intelligence tools that could provide a better understanding of targeted product or service patterns and trends, accurate market assessments, possible profit increase and early risk indications.

With the aim of improving data representation while trying to solve the variety challenge of the various data sources using sematic technologies, the construction of an enhanced business intelligence tool will be enabled that will provide better understanding of the industrial status and the financial market conditions leading to more accurate business assessments and better decisions.

Social media platforms represent one of the major Big Data sources, and have a great influence and impact on financial stock market and business investments [4, 20]. The financial domain thus represents a good area of investigation for the variety challenge of Big Data. This will be achieved by integrating structured data from the financial trade markets with the financial market news from Thomson Reuters (semi-structured to unstructured data), and with the semi-structured data from social media platforms such as Twitter. To achieve this aim, there is a need to use semantic ontologies such as Suggested Upper Merged Ontology (SUMO) and Financial Industry Business Ontology (FIBO).

In the next sections, introduction to different research components (Semantic Web, Big Data and Social Media) is provided followed by the challenges and the opportunities then the research position and selective research cases of similar work. Furthermore, the proposed research design and methodology are presented followed by the conclusion.

## II. DEFINITIONS FOR KEY TERMINOLOGIES

### A. Semantic Web

Semantic Web could be defined as a technology that allows the machines to understand the data available on the WWW and give responses that satisfy the users' requests more accurately and this is achieved by enriching and giving meanings to data using various kinds of web resources [21]. Semantic Web is extending users' abilities to share data beyond the obstacles and the limitations of programs and websites [22].

According to [22], the reasons behind the development of Semantic Web is to solve two issues: the limitations of data access (such as requesting documents using ambiguous terms in the query), and the delegation tasks' problems (for example the integration of different information) by providing the required access to data and enabling the delegation of specific types of tasks.

Semantic Web has been represented using many techniques and models such as those recommended by W3C as standard techniques including Extensible Markup Language, Resource Description Framework, and Web Ontology Language [21] where the last two are briefly explained below.

- The Resource Description Framework (RDF) is one of the very common semantic web languages. This kind of representation has the features to save resources' information via their own domain vocabularies [23]. The RDF records are constructed using three elements: resources (entities recognized by Uniform Resource Identifiers URIs), literals (atomics values such as numbers and text), and properties (connections identified by URIs) [22].
- The Web Ontology Language (OWL) is a representation technique that has more complex language and better machine-interpretability feature compared to RDF. It has the ability to define the resources' nature and their relationships more accurately [23]. This technique represents the

semantic data using ontology, "a formal representation of the concepts in some domain" [11].

### B. Big Data:

The continuous growth of complex data generated from various data sources makes traditional data analysis tools inadequate for processing and producing good results [15]. Big Data analysis can be defined as an alternative solution that has the ability to integrate, manage, and analyse heterogeneous and complex data type and sources [18]. Big Data is characterised by three main Vs (properties): Volume, Velocity, and Variety. Dealing with each one of these properties represents a major challenge that requires a good understanding of the issue and creative forms of information processing to bring more value from the data [24].

- Volume: Every day, around 2.5 quintillion bytes of raw data are generated from different data sources such as data streams, sensors, social media platforms, traffic and flight information systems, transaction records, and many others [15]. On one side, this massive size of data creates a huge obstacle for normal data analytics tools to process. On the other side, it creates a great opportunity to obtain valuable information that could improve the decision making of users and companies [15, 12].
- Velocity: Data is generated, transferred and processed at different speed, and this creates another challenge especially when dealing with different data streams in real time. This creates a challenge and a requirement for new models and algorithms that sufficiently process and analyse data streams in real time [18, 12].
- Variety: Data is generated from different sources using various formats and complexity. Mainly, data can be classified by three categories: (1) structured data such as relational database, (2) semi-structured data such as XML files, and (3) unstructured data such as text and social media posts. Trying to combine various data sources together creates a great challenge for traditional information system to handle. At the same time, successfully integrating and interoperating heterogeneous data creates a great potential to derive more insights and opportunities for better information systems [25].

In the recent years, Big Data analytical tools and applications have been used to solve complex problems in many fields and domains including global economy, marketing, national security, health, smart cities, business and commerce [18]. One example of the use of Big Data analytics is the e-commerce market transformation that is done by leading enterprises such as eBay and Amazon via adapting the Big Data analytics in their platforms and recommender systems [15]. Another example is the use of social media platform (such as Twitter) to analyse users' activities regarding marketing brands and to what extend social media influenced customers' perception and purchasing decisions which could have a positive effect in the overall companies' marketing strategies [15]. A last but

not least is the investment of Fortune 1000 companies in adapting Big Data technologies as critical important factor for firms to success and their current major driving of success is the integration of more and different sources [10].

## C. Social Media

Currently, social media is considered as one of the major Big Data sources that is generating huge amount of raw data via online users' activities and interactions [15]. Social media could be defined as an open digital environment that allows users to connect, interact, and collaborate with other users through different activities such as expressing their feelings, publish ideas, sharing opinions, and other daily activities [26, 15]. Popular social media platforms are including: Twitter, Facebook, LinkedIn, and Google+ [27].

Social media represents a very rich and valuable data source as it contains users' opinions, feelings, and interests embodied in their heterogeneous discussions and activities [15]. The wide spread of social media has a great impact on users' decisions and behaviours that could be for example noticed from brands rating, and place recommendations [28]. Text and sentiment analysis have been applied to this natural source of data streams for different reasons such as event detection, brand monitoring [28], trend analysis and financial stock market behaviour predication [20].

## III. CHALLENGES AND OPPORTUNITIES

The development and improvement in Big Data is still at early stages, and many challenges are unsolved and require further research. Examples of these challenges are: data access, data integration, data transmission, data analysis and data visualization [18, 15]. This research focus specifically on the data integration challenge, and detailed explanation is provided for this challenge specifically followed by the proposed solution to tackle this challenge in finance domain.

Managing distributed, large, and heterogeneous data from different sources is very challenging task for any information system [29, 18]. This could be related to various reasons such as data collection and maintenance actions that change the data form, the changes in data structure over the time, the lack of central data management tasks which leads to the creation separate subsystems losing the main target of having one coherent system [30, 7].

A number of systems have adapted the use of service-oriented approach and data warehousing techniques [30]. However, these systems have failed in solving the data integration challenge and they even added more complexity to the problem. For these reasons, organisations and large companies are investing money and time to create a system that unified different data sources in effective way [30].

Semantic ontology based data access (OBDA) approach has been proposed by researchers, that uses semantic web and ontology techniques, to tackle the data integration challenge. By using this approach, a layer that mediated different data sources will be created, which allows the access and query all connected data using single access point [7]. This kind of approach, that uses semantic technology, has a great advantages and potentials in solving the variety of data sources and dealing with other constraints such as large data volume, and velocity [12].

Examples of possible benefits are including: (1) the re-usability of extracted knowledge is achieved as the domain is represented explicitly; (2) the use of ontology and the mapping allows easy data governance and management; (3) the flexibility and extensibility of such a system are important features that allow incremental integration of available and new data sources [30]; and (4) the minimizing of required changes in exist data management infrastructure [7].

## IV. POSITIONING OF THE RESEARCH

In recent years, the huge amount of data created and collected has exceeded the human ability to analyse it and to extract hidden and useful patterns and knowledge. Developments in computational power, data structure and algorithms allow Big Data solutions to materialize and handle the large amount of data, offering the capability to disclose hidden insights from the abundant data [3, 1].

Many companies are supporting and developing Big Data technology and analytics for example Google, SAS, and IBM [31], providing Big Data solutions such as Apache Hadoop, IBM BigInsights, and IBM Watson.

Today, the financial industries and organizations (small or large) are facing difficulties in data integration and management, dealing with the heterogeneity of data formats and meaning, and the volume of data generated from different information systems [32].

Currently, possible developed semantic ontologies that could be adapted include Financial Industry Business Ontology FIBO [17,33] and Suggested Upper Merged Ontology (SUMO) [34] in finance domain, and Friend-of-a-Friend (FOAF) and Semantically Interlinked Online Communities (SIOC) in online networking domain [35, 6].

Another interesting component of the proposed research is the social media platform, for example, Facebook and Twitter, which are among the most interesting and rich sources of data, as they reflect the users' opinions, interests and feelings [15, 36]. These communities generate a large amount of data every day: Twitter has more than 280 million monthly active users who exchange more than 500 million messages (tweets) daily [6].

The challenging task in a social media platform (such as Twitter) is to identify, extract and analyse information from such a heterogeneous, large-scale and noisy data source. This requires a combination of different domains and knowledge to achieve the required goals [29, 19]. Conceptually, the combination of social media, Big Data and data analysis could be represented as shown in Fig. 1 [15].

According to [37], semantic ontology is representing a promising approach to solve complex problems and suggest it to detect the financial stock market manipulators. The research specifically focuses on the trade-based manipulators, large or group of traders who try to change the assets' value via the stock market buying and selling processes. The work is done by design and constructs three ontologies: static, dynamic and social ontologies and demonstrate the effectiveness of the approach on Zhong Ke

Manipulation case study, one of the biggest stock market manipulation cases in China. The work is missing the information about the construction and the representation of the created ontologies and more information is required on how the evaluation and the analysis have been carried out.
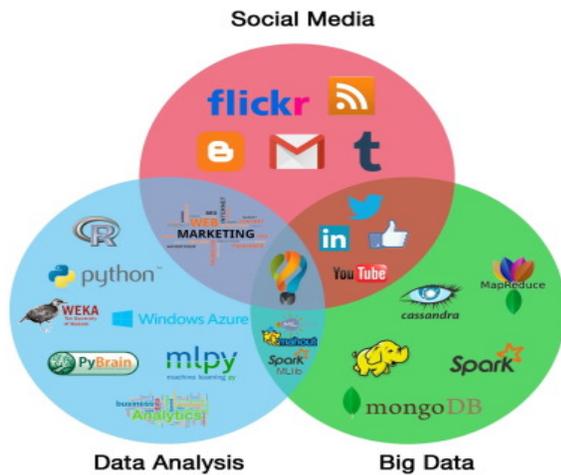


Figure 1. The conceptual representation of social Big Data (taken from [15]).

A search engine that is especially designed to deal with financial news is proposed in [38]. They suggest the use of semantic technologies as the main approach to handle the heterogeneity of financial-related news. The work is carried out by constructing financial ontology using OWL representation and utilising it to annotate RSS feeds from different financial sources and create populated data (instance data that match ontology structure). The validity of the approach is demonstrated via a case study within the Spanish stock market exchange.

Another line of research [39] notes that the tone and the sentiment words in financial news articles have a direct impact on the financial stock market and analysing them could provide an indication of the stock market price movements. By using sentiment analysis and machine learning techniques, a financial prediction system is implemented and a test is carried out using one month of collected news articles. The approach shows good results regarding the market prediction. However, the proposed system does not use semantic technology to deal with data which is unstructured, but instead they use a binary matrix to represent the existence of corpus' terms in articles.

A case study presented by [17], mentions the sentimental and emotional analysis of social networks in the financial domain using a combination of semantic financial ontologies (FIBO), social networks (Twitter) and other resources that provide uniform vocabulary to express sentiments and emotions in a proper format (Marl, and Onyx). Analysing the gathered opinions, emotions, and activities connected to services, brands and products from Twitter with the FIBO and other data resources gives better insights into the investment mood of different communities, which could be an effective factor in the financial domain. The work is implemented in a proper and organized way and the proposed system is explained in detail. However, the approach requires further effort. There is a lack of evaluation and validation of the proposed techniques, and the conclusion is general and does not clearly reflect the benefits of the proposed approach.

The work presented by [2], examines the use of semantic modelling techniques to represent knowledge on the social media community (such as Twitter) to be used as a basis for recommendation systems. The proposed approach is to construct a semantic ontology model using content representation from social media, existing vocabularies and sentiment analysis and help from experts from the target domain. This work has been validated via an experiment carried out on Twitter (as a social media data source), using specific brands of vehicle, and with an expert in the automotive industry domain to define the rules and characteristics of the products in the created ontology. The case study reveals interesting results that validate the proposed concept, linking users with the brands in the targeted period of time, yet it does so regardless of the user opinions of the target brands (sentiment scores and rates).

Another proposal [4] claims that the direction of movement of stock market prices can be predicted by analysing the social media represented by Twitter and news from financial markets. The proposed approach is to analyse social media messages and extract their sentiments to help increase the prediction of the stock market direction. The Multiple Kernel Learning – Support Vector Machine (MKL-SVM) is used to process both news articles and tweet sentiments. The research shows an improvement in the accuracy of the prediction, yet the analysis does not use semantic technology .

According to [40], a change is taking place in the financial services domain, where the main focus is on the commoditizing of data. The work concentrates on moving from traditional data management (such as Extract-Transform-Load (ETL) technologies) to a new approach supporting semantic technologies, which will give the opportunity to support dynamic data formats and structures and is scalable to encounter upcoming and new consumption purposes. The work is done by modifying the data ingestion tier and using the Big Data storage tier combined with semantic ontologies (such as FIBO) and semantic query tools, followed by insights and a visualization tier (which has the ability to combine multi-dimensional queries from different sources). The work has a shortage of information regarding the implementation, and no evaluation and validation of the proposed system.

A case study presented by [8] proposes a new system, called Karma, that uses semantic technology to integrate different type of Big Data sources trying to solve the variety challenge. A detailed information of the various processing steps have been provided starting with importing different type of data sources, cleaning process, modelling and its problems (nomenclature and format differences), ending with the integration process. The proposed system is supporting only specific types of structural and semi-structural data sources. It uses semantic RDF technology to tackle integration issues such as nomenclature and

identifying same entities in different datasets at schema level. The work has a good explanation to the integration problem, and the process to solve it. However, the work is limited to specific types of data sources and does not include unstructured data sources in their considerations.

The work presented by [24] represents a case study in the health domain that is applying similar approach to the proposed one. The work carried out to integrate different data sources from the medical and oral health domain using semantic ontology to unite them and allowing the construction of evidence-base system that provides more comprehensive decision support to healthcare professionals regarding patient's situation. To apply the approach, different processing steps have been done including: collecting domain information, constructing the domain ontology, developing the rules, and implementing and evaluating the system. The research shows interesting results proving the validity of semantic approach in solving the data integration and handling the variety challenge.

An event detection and trend analysis system is proposed in [6]. The system uses Twitter data to create an ontology module, which is one of the main components of the proposed system. Different components have been used to construct the required system, including a tweet parser and entity extractors, among others. The testing is carried out by launching two smartphones onto the market and the system reveals interesting results in detecting users' activities regarding these products. The research requires more attention in describing the system details and how the system's components are used to complement one another. The proposed system does not consider the sentiments in the tweets and the construction of the ontology is limited to the top 100 keywords extracted from the collected tweets. More work is required to improve the system and in the creation of the ontology.

The presented cases revealed the recent research on using semantic ontologies to integrate different data sources and deal with structured and unstructured data (summarised in Table I). Various approaches and data sources are used trying to achieve better results and/or tools solving the issues of complex formats, structures and meanings. Yet, the majority of the works used limited and basic semantic ontologies and did not fully imply the Big Data for example [35, 40]. Currently, the enhancement of Big Data using ontologies and semantic techniques to solve the variety challenge is at the early stages and still a research challenge that requires further work especially in the business and financial domain, which creates an opportunity to have a good contribution to the knowledge.

Utilizing semantic technologies to empower Big Data solutions could improve and enrich system solutions, especially in the business domain solving difficulties facing customers [41]. The combination of Big Data solutions and semantic technologies is still new and more work is required in improving the semantic Big Data representation, data processing and analysis [16], and in creating an effective platform that could provide a better understanding of the targeted trade market and the influence of social media.

## V. RESEARCH DESIGN AND METHODOLOGY

The Design Science Research (DSR) Process Model, shown in Fig. 2 [42], is suggested to carry out the proposed research.
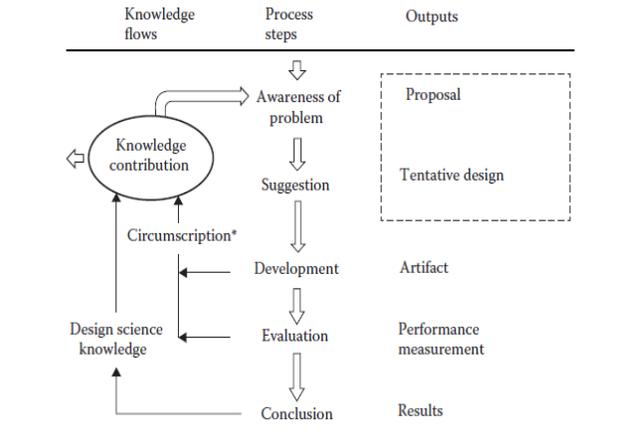


* Circumscription is discovery of constraint knowledge about theories gained through detection and analysis of contradictions when things do not work according to theory (McCarthy, 1980)

Figure 2. The DSR Process Model (taken from [42]).

### A. Awareness of the problem

The actions of industries, organizations, and people are reflected in strategy, processes, services, and products. These actions are built on gathered data, analytics and the decisions of business industries. This reveals the urgent need for approach which can extract semantic knowledge from Big Data, and to allow business users and decision-makers to use the full potential of available data effectively to uncover insights that could drive innovations, profits, and success. This means that data analysis is required to improve and incorporate the new mechanisms for semantic enhancement of Big Data and how it effectively and confidently gives accurate decisions to decision-makers.

TABLE I.        SUMMARY OF THE PRESENTED CASES.

| Reference No | Research Domain | Method | Types of data sources | Data Sources | Ontology Construction Type | Notes |
|---|---|---|---|---|---|---|
| [37] | Financial stock market | Semantic ontology representation | | | Constructed | Using ontology representation to better understanding the stock market manipulation |
| [38] | Financial news | Search engine using semantic ontology | Semi-structured | RSS feeds + Spanish stock | Constructed | Creating search engine using semantic technology |

| Reference No | Research Domain | Method | Types of data sources | Data Sources | Ontology Construction Type | Notes |
|---|---|---|---|---|---|---|
| | | | | exchange | | |
| [39] | Financial news | Sentiment analysis and machine learning | Unstructured | Financial news | | Detecting the financial stock market price from the financial news articles |
| [17] | Social media and financial | Sentiment analysis using semantic ontology | Unstructured | Twitter | Exist - FIBO | Detecting the investment mood via social media |
| [2] | Social media and car industry | Sentiment analysis and semantic modeling | Unstructured | Twitter | Constructed | Using social media to improve the recommendation system |
| [4] | Social media and financial | Sentiment analysis and predication analysis | Unstructured and structured | Twitter and daily stock market | | Using social media sentiment analysis to predicate the stock market predication |
| [40] | Financial | Data aggregation using semantic technologies | | | Exist - FIBO | Shortage of information regarding the financial data services and the applying and implementation of the proposed approach |
| [8] | Art and museum | Data integration using semantic technologies | Structured and semi-structured | Crystal Bridges Museum | Constructed by reusing exist ones | The work does not include unstructured data sources. |
| [24] | Medical | Data integration using semantic technologies | Structured, unstructured and semi-structured | Medical and oral health condition | Constructed by reusing exist ones | |
| [6] | Social media | Event detection and trend analysis | Unstructured | Twitter | Constructed by reusing exist ones | |

The proposed research questions (artifacts) are: How can semantic technologies be used to resolve the variety challenge of the Big Data sources? How can semantic knowledge be extracted from the Big Data sources?

*B. Suggestions*

The main design of the proposal is to work on constructing an approach that combines the semantic technologies and Big Data, using social media and financial news and financial trade markets as main data sources as shown in Fig. 3.
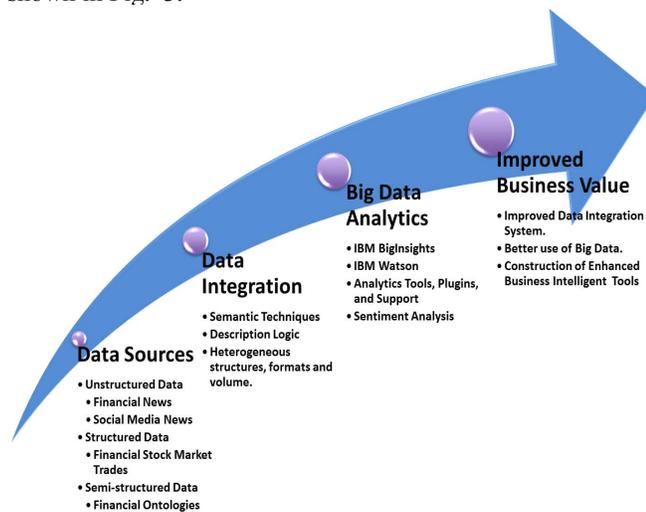


Figure 3. The elements and the steps of the proposed approach.

Regarding Big Data, we will analyse the available data sources and Big Data tools and applications (such as IBM Watson, IBM BigInsights and Apache Hadoop) and how we could combine this with the semantic data sets. The search will be conducted keeping in mind the availability of semantic ontologies such as FIBO, FOAF and SIOC and how we could adapt them to construct new ontology and integrate it with the existing Big Data application.

*C. Development and evaluation*

By mastering the science of understanding the meaning of exploring Big Data, a new model will be created that could be significant. This model will deal with the semantic extraction of Big Data sources and create the required data integration that allows data analysis to be carried out. In the analysis part, the existing methods and techniques might need to be enhanced, modified, or re-engineered to work with the proposed approach. Testing, evaluation and validation of the proposed approach will be confirmed by developing the prototype concept based on the available tools, which will illustrate the advantage of this approach.

VI. CONCLUSION

As we live in the digital age, "Data is the new oil in the digital economy" [43]. It is the most important factor in our world and is reflected in every action we take (buying product and bank transactions, among others), which creates the Big Data world. The proposed approach will help in improving the use of data, enabling the construction of better intelligent data analytics/business intelligence tool.

REFERENCES

[1] T. Kraska, "Finding the needle in the Big Data systems haystack," IEEE Internet Computing, vol. 17(1), pp. 84-86, 2013, doi: 10.1109/MIC.2013.10.

[2] D. Villanueva, I. González-Carrasco, J.L. López-Cuadrado, and N. Lado, "SMORE: Towards a semantic modeling for knowledge representation on social media," Science of Computer Programming, 2015, doi:10.1016/j.scico.2015.06.008.

[3] S. Sagiroglu, and D. Sinanc, "Big Data: a review," Proceedings of the Collaboration Technologies and Systems (CTS), San Diego, CA. 20th - 24th May 2013, pp. 42-47, 2013, doi: 10.1109/CTS.2013.6567202.

[4] K.N., Devi, and V.M., Bhaskaran, "Semantic enhanced social media sentiments for stock market prediction," World Academy of Science, Engineering and Technology, International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering, vol. 9(2), pp.684-688, 2015.

[5] K. Guo, R. Zhang, and L. Kuang, "TMR: Towards an efficient semantic-based heterogeneous transportation media Big Data retrieval," Neurocomputing, vol. 181, pp.122-131, 2016.

[6] R. Kaushik, S.A. Chandra, D. Mallya, J.N.V.K. Chaitanya, and S.S. Kamath, "Sociopedia: an interactive system for event detection and trend analysis for twitter data," Proceedings of 3rd International Conference on Advanced Computing, Networking and Informatics (ICACNI-2015), KIIT University, India. 23th – 25th June 2015. Springer India, pp. 63-70, 2016.

[7] E. Kharlamov et al., "Ontology based access to exploration data at Statoil," Proceedings of the 14th International Semantic Web Conference (ISWC), Bethlehem, Pennsylvania, 11th – 15th October 2015. Springer International Publishing, pp. 93-112, 2015.

[8] C.A. Knoblock, and P. Szekely, "Exploiting semantics for Big Data integration," AI Magazine, vol. 36(1), pp. 25-39, 2015.

[9] R. Bean, "Variety, Not Volume, Is Driving Big Data Initiatives," MITSloan Management Review, 28 March, 2016. Available at: http://sloanreview.mit.edu/article/variety-not-volume-is-driving-big-data-initiatives/ (Accessed: 17 April 2017).

[10] NewVantage Partners, "Big Data Executive Survey 2016: an update on the adoption of Big Data in the Fortune 1000, Executive summary", 2016. Available at: http://newvantage.com/wp-content/uploads/2016/01/Big-Data-Executive-Survey-2016-Findings-FINAL.pdf (Accessed: 25 April, 2017).

[11] M. Atkin, and M. Bennett, "Semantics in systemic risk management," in J. Fouque, and J. Langsam, Eds., Handbook on Systemic Risk. New York: Cambridge University Press, 2013, pp. 123-160.

[12] K. Thirunarayan, and A. Sheth, "Semantics-empowered approaches to Big Data processing for physical-cyber-social applications," Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI), Symposium on Semantics for Big Data, Arlington, Virginia, 15th - 17th November 2013, pp. 68-75, 2013.

[13] E. Hillerbrand, "Semantic Web and business: reaching a tipping point?, in M. Workman, Ed., Semantic Web. London: Springer International Publishing, 2016, pp. 213-229.

[14] G. Vugrala, G. Bafna, and S. Pawar, "Semantics and ontology – the future of data aggregation," Wipro, 2015. Available at: http://www.wipro.com/documents/semantics-and-ontology-the-future-of-data-aggregation.pdf (Accessed: 22 April, 2017).

[15] G. Bello-Orgaz, J.J. Jung, and D. Camacho, "Social Big Data: Recent achievements and new challenges," Information Fusion, vol. 28, pp.45-59, 2016.

[16] A. Prasad, and L. Feigenbaum, "How Semantic Web Tech Can Make Big Data Smarter," 2014. Available at: http://www.cmswire.com/cms/featured-articles/how-semantic-web-tech-can-make-big-data-smarter-026726.php , (Accessed: 22 April, 2017).

[17] 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.

[18] F.E. Jamiy, A. Daif, M. Azouazi, and A. Marzak, "The potential and challenges of Big Data-Recommendation systems next level application," International Journal of Computer Science Issues, vol. 11(5), pp. 21-26, 2014.

[19] K. Bontcheva, and D. Rout, "Making sense of social media streams through semantics: a survey," Semantic Web, vol. 5(5), pp.373-403, 2014.

[20] B.D. Egger, "Social media strategies for investing: how twitter and crowdsourcing tools can make you a smarter investor," Avon, MA: Adams Media, 2014.

[21] D. Jeon, and W. Kim, "Development of semantic decision tree," Proceedings of the 3rd International Conference on Data Mining and Intelligent Information Technology Applications (ICMiA). IEEE Conference Publications, pp. 28-34, 2011.

[22] J. Domingue, D. Fensel, and J.A. Hendler, Handbook of Semantic Web technologies, Eds. Heidelberg, Berlin: Springer-Verlag, 2011.

[23] V. Sugumaran, and J.A. Gulla, Applied Semantic Web technologies, Eds. Boca Raton, FL: Taylor & Francis Group, 2012.

[24] T. Sheh, F. Rabhi, and P. Ray, "Investigating an ontology-based approach for Big Data analysis of inter-dependent medical and oral health conditions," Cluster Computing, vol. 18(1), pp. 351-367, 2015, doi: 10.1007/s10586-014-0406-8.

[25] R.T. Bedeley, and L.S. Iyer, "Big Data opportunities and challenges: the case of banking industry," Proceedings of the Southern Association for Information Systems Conference (SAIS), Macon, GA. 21th-22th March 2014. pp. 1-6, 2014.

[26] P. Burnap, et al., "COSMOS: Towards an integrated and scalable service for analysing social media on demand," International Journal of Parallel, Emergent and Distributed Systems, vol. 30(2), pp.80-100, 2015.

[27] A. Van Looy, Social media management: technologies and strategies for creating business value. Heidelberg, NY: Springer, 2016.

[28] C. Musto, G. Semeraro, P. Lops, and M. de Gemmis, "CrowdPulse: a framework for real-time semantic analysis of social streams," Information Systems, vol. 54, pp.127-146, 2015.

[29] U. Sivarajah, M. M. Kamal, Z. Irani, and V. Weerakkody, "Critical analysis of Big Data challenges and analytical methods," Journal of Business Research, vol. 70, 263-286, 2017.

[30] C. Daraio, M. Lenzerini, C. Leporelli, P. Naggar, A. Bonaccorsi, and A. Bartolucci, "The advantages of an Ontology-Based Data Management approach: openness, interoperability and data quality," Scientometrics, pp. 1-15, 2016.

[31] J. Hurwitz, A. Nugent, F. Halper, and M. Kaufman, Big Data for dummies. Hoboken, NJ: John Wiley & Sons, 2013.

[32] T. Butler, and E. Abi-Lahoud, "A mechanism-based explanation of the institutionalization of semantic technologies in the financial industry," in B. Bergvall-Kåreborn, and P.A. Nielsen, Eds., Creating value for all through IT. New York: Springer Berlin Heidelberg, pp. 277-294, 2014.

[33] EDM Council, "FIBO Semantics Repository", 2015. Available at: http://www.edmcouncil.org/semanticsrepository/index.html , (Accessed: 22 April, 2017).

[34] A. Pease, Suggested Upper Merged Ontology (SUMO), 2016. Available at: http://www.adampease.org/OP , (Accessed: 22 April, 2017).

[35] L.A. Cotfas C. Delcea, I. Roxin, and R. Paun, "Twitter ontology-driven sentiment analysis," in D. Barbucha, N.T. Nguyen, and J. Batubara, Eds., New trends in intelligent information and database systems. London: Springer International Publishing, 2015, pp. 131-139.

[36] H. Watson, K. Wadhwa, L. Baruh, and S. Scifo, "Social networking in times of crisis," in A. Kurylo, and T. Dumova, Eds., Social

networking: redefining communication in the digital age. London: Rowman & Littlefield, 2016, pp. 111-130.

[37] L. Siming, and W. Huaiqing, "Ontologis for Stock Market Manipulation," Proceedings of the 2011 International Conference on Informatics, Cybernetics, and Computer Engineering (ICCE2011), Melbourne, Australia, 19th – 20th November 2011, Springer Berlin Heidelberg, pp. 1-9, 2011.

[38] E. Lupiani-Ruiz, et al., "Financial news semantic search engine," Expert systems with applications, vol. 38(12), pp. 15565-15572, 2011.

[39] R.P. Schumaker, Y. Zhang, C.N. Huang, and H. Chen, "Evaluating sentiment in financial news articles," Decision Support Systems, vol. 53(3), pp. 458-464, 2012.

[40] S. Gollapudi, "Aggregating financial services data without assumptions: a semantic data reference architecture," Proceedings of

Semantic Computing (ICSC), Anaheim, CA. 7th – 9th February 2015. IEEE International Conference, pp. 312-315, 2015. Doi: 10.1109/ICOSC.2015.7050825.

[41] IBM, "Using semantic enrichment to enhance Big Data solutions," 2012. Available at: http://www-01.ibm.com/software/ebusiness/jstart/semantic/ , (Accessed: 10 February 2016).

[42] V.K. Vaishnavi, and W. Kuechler, Design science research methods and patterns: innovating information and communication technology, CRC Press, 2015.

[43] S.V. Shankar, "Data is the new oil of the Digital Economy," 2015. Available at: http://bigdata-madesimple.com/data-is-the-new-oil-of-the-digital-economy/, (Accessed: 27 April, 2017).