

A Web-Based Framework for Classification and Integration of News Articles Big Data using Ontology



Jeelani Ahmed, Muqem Ahmed

Abstract: Around 2.5 quintillion bytes of data have been created online: out of which most of the data has been generated in the last two years. To generate this huge amount of data from different sources, many devices are being utilized such as sensors to get the data about climate information, social networking sites, banking records, e-commerce records, etc. This data is known as Big Data. It mainly consists of three 3v's volumes, velocity, and variety. Variety of data discusses about different formats of data originating from various data foundations. Hence, the big data variety's issue is significant in explaining some genuine challenges. The semantic Web is utilized as an Integrator to join information from different sorts of data foundations like web services, social databases, and spreadsheets and so on and in various formats. The semantic Web is an all-encompassing type of the present web that gives simpler methods to look, reuse, join and offer the data. In this manner, it is along these lines seen as a combiner transversely over different things, information applications, and systems. This paper is an effort to uncover the nature of big data and a brief survey on the use of various semantic web-based methods and tools to add value to today's big data. In addition, it discusses a case study on performing various machine learning functionalities on news articles and proposes a web-based framework for classification and integration of news articles big data using ontologies.

Keyword: bigdata, semantic web, ontology, data integration, news article

I. INTRODUCTION

The big data term was instituted to speak to the vast sum and numerous kinds of advanced information that we are using today, including records, pictures, recordings, sound, and sites [1]. Semantics-based methodologies are viewed as helpful methods for managing vast scale information, for example, big data. So as to investigate this theme, it is first important to all the more unmistakable portraying the idea of enormous information.

A. Big Data

The present world is being digitized quickly and various sorts of colossal measure of information are being created from different sorts of heterogeneous information sources, which influence the various elements of the common public like individuals, association and businesses in a different way, for instance, different internet business exchanges and rating administrations, banking and ATM exchanges [2],

clients mentality through different online networking stages, nonstop sound video accounts, and sensor-produced information. What's more, to use this enormous measure of crude information for advantages of the business, we should need to comprehend the information that is vital for tracking activities that could build the item deals, recognize the different patterns in the market and diminishes the business misfortunes and in this manner making more choices accessible for the chiefs of the business [1] [3].

Big data is characterized as comprising principle 3v's: volume, variety, and velocity [4]. The volume discusses the gigantic measure of the information produced through various gadgets and how to oversee huge instructive accumulations, which ordinarily requires execution in a pass on the cloud-based establishment. Velocity implies overseeing constant spouting data, for instance, the video continues, where it may be hard to store all data for later getting ready. Variety implies overseeing various sorts of sources, particularly game plans of the data, and sweeping amounts of sources [4]. A critical piece of the work of tremendous data has focused on volume and Velocity; in any case, the issues of Variety are likewise indispensable in dealing with various genuine issues.

- **Volume:** On each minute immense quantity of information is generated from numerous data sources like banking dealings, records, social media networks, continuous knowledge streams, traffic systems, instructional resources, flight data systems sensors and far a lot of. Old data analysis tools facing difficulties to method and analyze this immense quantity of complicated data therefore there by trendy knowledge analysis tools want to method this large data.

- **Velocity:** Enormous Data is being produced by some continuous sources, for example, information streams, sensors and live spilling of various information accounts. These information age sources produce the information at various rates; again, information can be handled and exchanged at various velocities. Along these lines of information, age offers a test to deal with the information, particularly when information is coming continuously from a wide range of streams.

Manuscript published on 30 September 2019

* Correspondence Author

Jeelani Ahmed*, Department of CS & IT, Maulana Azad National Urdu University, Hyderabad, India. Email: jeelani.jk@gmail.com

Muqem Ahmed, Department of CS & IT, Maulana Azad National Urdu University, Hyderabad, India. Email: Muqem.ahmed@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

• **Variety:** Variety states that data generated in different formats from different sources. Usually data may be structured, semi-structured or unstructured data for example relational databases, XML pages and social media posts respectively. Dealing with a variety of data is a big challenge as it offers several type data to be processed and to be integrated into one format. Semantic web plays an important role in dealing with the variety challenge of big data.

In a 2001 Scientific American article by Tim Berners-Lee (creator of the World Wide Web), and co-authors James Hendler and Ora Lassila [5] demonstrate the advanced sight of the web, where data are interconnected in a meaningful manner. Reasoning and filtering of data and ontology; these are the two main rivulets that have emerged from the semantic web. SPARQL (SPARQL Protocol and RDF Query Language), Resource Description Framework (RDF) and OWL (Web Ontology Language) these are the currently exists standards of semantic web. These standards make information access available for the end-users when they require at the time. The World Wide Web considers the semantic web as an extension as opposed to a substitution of the present web. As talked about over, one of the main concepts of the semantic web is ontology, which comprises ontology evolution, engineering, merging, matching and mapping. Another standard, which comprises filtering and reasoning, which contains the area of collaborative filtering, content filtering, reasoning and hybrid filtering.

II. SEMANTIC CHALLENGES IN HANDLING BIG DATA

There are numerous challenges in working with big data in the sense of volume, variety and velocity, which comprises capture, storage [6], access [7], sharing, search, analysis [8], visualization and transfer. However, following areas present the semantic challenges in working with big data:

A. Capturing large scale big data

One profoundly appropriate and huge test is the progressive auto-course of action of unstructured information into arranged ideas. This ought to include getting information and related ideas and effectively doling out the information to the correct ideas and classes without including human/manual analysis [9]. This test identifies with the variety and volume of huge information that is produced by the web, internet based life, gushing, web journals, and different sources. The auto-plan of data as directed by ideas requires a comprehension of settings and substance. This ought to be founded on an examination of new information with a current learning base, just as the advancement of scientific categorizations. Because of enormous information, there could be billions of triples connected to one another. The preparing of such an extent of information must include effective ontological designing methodologies.

B. Getting the relevant and right data from big data

Revealing and getting to the applicable information inside the huge volume and variety of huge information is an

inexorably troublesome errand [7] [9]. As indicated by [10], issues emerge because of the three elements of huge information: volume, since huge measures of information have been aggregated throughout the decades, volume, since the sums might be quickly expanding, and variety, since the information is spread over various arrangements. While this is an ordinary issue of managing enormous information, applications can be utilized for auto-comprehend the elements of the setting. For example, traffic information can be added to guides to give setting on street conditions, the likelihood of postponement, length of anticipated impediments, street situations, and so on.

C. Linking big data

The concept of linking data is popular and has gained much importance across the world and many companies such as Google and Facebook has been adopted successfully [11].

III. EXISTING SEMANTIC-BASED TOOLS FOR BIG DATA

Table 3 presents a summary of semantics-based tools for dealing with big data. Our research only discovered a limited number of such tools. A number of non-commercial tools are still ongoing projects, and many focus on specific domains. For example, SINA [29] focuses on the medical domain.

Table I: Semantic/ontology-based tools for big data

Tool	Features Offered	Semantics
Oracle Big Data Appliance [23]	<ul style="list-style-type: none"> • Hadoop Distributed File System [24] • Oracle NoSQL Database • Storage capacity of 648TB 1152GB of memory 	<ul style="list-style-type: none"> • Oracle Exadata • Oracle database • Integrated management
IBM Watson Foundations [25]	<ul style="list-style-type: none"> • Hadoop Distributed File System [24] • IBM InfoSphere Platform IBM Stream Computing 	<ul style="list-style-type: none"> • Mapping • Stream computing
Ontology 4 Platform [26]	<ul style="list-style-type: none"> • Semantic search across an enterprise • Structured and unstructured data handling 	<ul style="list-style-type: none"> • Ontology • Semantic search
Optique: OBDA Solution [7]	<ul style="list-style-type: none"> • EU FP7-funded Optique, which develops a start to finish OBDA framework, giving adaptable end-client access to advanced big data stores 	<ul style="list-style-type: none"> • Ontologies • OWL 2
PigSPARQL [27]	<ul style="list-style-type: none"> • Utilizes Pig (Latin) and it is built on Hadoop 	<ul style="list-style-type: none"> • SPARQL
SINA [28]	<ul style="list-style-type: none"> • Three integrated medical databases • Natural language support 	<ul style="list-style-type: none"> • RDF • Linked data



Treo [29]	<ul style="list-style-type: none"> Heterogeneous databases are supported. Thorough semantic coordinating dependent on distributional semantics Linked Data/RDF supports queries 	<ul style="list-style-type: none"> RDF Linked data 	spreadsheets.	
BioPortal [30]	<ul style="list-style-type: none"> Web services integrated into software applications 	<ul style="list-style-type: none"> Ontologies Semantic mapping 		<ul style="list-style-type: none"> SPARQL RDF
DIVE [31]	<ul style="list-style-type: none"> Graph-based visual analytics 	<ul style="list-style-type: none"> Ontology Direct SQL 		
KRMA [32]	<ul style="list-style-type: none"> Data from various sources/databases and in different formats can be integrated like text files, JSON, KML and 	<ul style="list-style-type: none"> RDF 		
			LODatio [33]	<ul style="list-style-type: none"> Semantic search engine Schema-based record to distinguish significant hotspots for LOD for given SPARQL query designs.
			SchemEX [34]	<ul style="list-style-type: none"> Linked data indexing based on stream.

IV. EXISTING METHODS TO THE SEMANTIC COMPUTING OF BIG DATA

Many methods have been proposed for semantic approaches for big data. These proposed methods work at various stages and consists of areas such as ontology evolution, ontology engineering, and reasoning, representing big data and matching. Such approaches are

Table II: Semantics approaches for big data

Reference	Issue	Approach	Main Focus	Variables / Parameters
[13]	Utilizing reasoning to improved query results over an extremely gigantic measure of information (i.e., Web-scale information), based on a parallel and conveyed framework.	<ul style="list-style-type: none"> Map Reduce Reasoning algorithms Grouping RDFS rules in four Map Reduce jobs. 	<ul style="list-style-type: none"> Web-scale reasoning 	Volume: One billion to 100 billion RDF triples
[14]	Providing access to big data	<ul style="list-style-type: none"> Ontological-based techniques, included by query optimization and parallelization Iterative user-centric development approach 	<ul style="list-style-type: none"> Visual query formulation for the end-user, demonstrating the introductory visual query system based on ontology. Alleviating the effects of big data 	Data dimensions such as velocity, volume and variety and also schemas.
[15]	Presenting contents displayed on a page dynamically, on the basis of viewer's circumstances	<ul style="list-style-type: none"> New technique for user profiling, on the basis of information from user surfing logs User Characteristics must probably show by the user profiles. Presents relations between creating components and the capriciousness that comes from the mechanized handling of genuine information. 	<ul style="list-style-type: none"> Using Ontology engineering, data analysis and resource processing of big data for tackling the issue. For each user in the system, profile ontology has been automatically constructed and populated. 	

[16]	Within the Optique OBDA system Query optimization and transformation with respect to query answering.	<ul style="list-style-type: none"> • Queries presented in heterogeneous distributed databases and streams for automatic big data query generation 	<ul style="list-style-type: none"> • Within the Optique OBDA system Query optimization and transformation with respect to query answering. • Two systems Quest and Pegasus who supports Optique's automatic generation of queries. 	User-friendly query formulation framework, Ontology maintenance, mapping, analytics and processing over streaming data, distributed query optimization,
[17]	Stream reasoning	<ul style="list-style-type: none"> • Analysis of related research fields, with the aim of extracting algorithms, models, techniques, and solutions 	<ul style="list-style-type: none"> • Identifying the requirements involved in various application setups and isolating the issues that they pretend. • Examines the previous methods and applications in the stream reasoning and emphasizing their strong points and weakness. 	Frequent change of huge information
[18]	Semantic inconsistencies between applications and systems	<ul style="list-style-type: none"> • Management of metadata in smart grids to remove semantic inconsistencies 	<ul style="list-style-type: none"> • An Ontological model has been developed for offshore wind energy meta-data management 	Data exchange and Knowledge sharing, Obtained information on relationships between theories, Data value as metadata
[19]	Use of FIBO (financial industry business ontology) as an abstract model.	<ul style="list-style-type: none"> • FIBO (Financial industry business ontology) 	<ul style="list-style-type: none"> • Possible creation of such applications based on semantic technology which can be used to perform novel types of records processing. 	FIBO offers financial business ontology

Table III: Ontology evolutionary approaches for big data

Reference	Issue	Approach	Main Focus	Variables / Parameters
[14]	Providing access to scalable big data	<ul style="list-style-type: none"> • Ontological-based approach with query optimization and parallelization. • Iterative user-centric development approach 	<ul style="list-style-type: none"> • Visual query formulation for the end-user, demonstrating the introductory visual query system based on ontology. • Alleviating the effects of big data 	Data dimensions such as velocity, volume and variety and also schemas.
[20]	Solution for big data integration and scalability Ontology-based data access (OBDA) has been provided in which using a similar concept of the fundamental domain end-users formulate the queries.	<ul style="list-style-type: none"> • Collection of likely modeling and logical errors in OBDA frameworks and the primary difficulties looked at supporting the life-cycles of OBDA frameworks. 	<ul style="list-style-type: none"> • Ontology maintenance and mapping problem, bootstrapping problem. • Main test of correcting bugs in ontologies and mappings. • Describing various instances of types of errors, and offering primary views on how to correct them. 	Creation, preservation, and alteration of an OBDA specification.

Reference	Issue	Approach	Main Focus	Variables / Parameters
[21]	Direct entry of information not presently out there on the online or sure up in machine-readable text archives	<ul style="list-style-type: none"> • Consumption and Publication of interconnected data. • Drawing on the practical linked data scenario • Choosing what information to return in a depiction of an asset on the web. • Techniques and systems for computerized connecting of informational collections. 	<ul style="list-style-type: none"> • The study of previous interconnected data systems and architectures. 	
[22]	Ontology matching evaluations	<ul style="list-style-type: none"> • Experimental review of matching systems 	<ul style="list-style-type: none"> • SAMBO (Linköping U.) • Agreement Maker (U. Of Illinois at Chicago) • Falcon (Southeast U.) • Anchor-Flood (Toyohashi U. of Technology) • DSSim (Open U., Poznan U. of Economics) • ASMOV (INFOTECH Soft, Inc., U. of Miami) • RiMOM (Tsinghua U., Hong Kong U. of Science and Technology) 	Semantic technologies for ontology matching

V. USE CASE: MAP NEWS DATA INTEGRATION

The motivation for our use case comes from integrating the news articles from various sources with location data using ontologies. The proposed framework for the use case shown in the figure 5.1

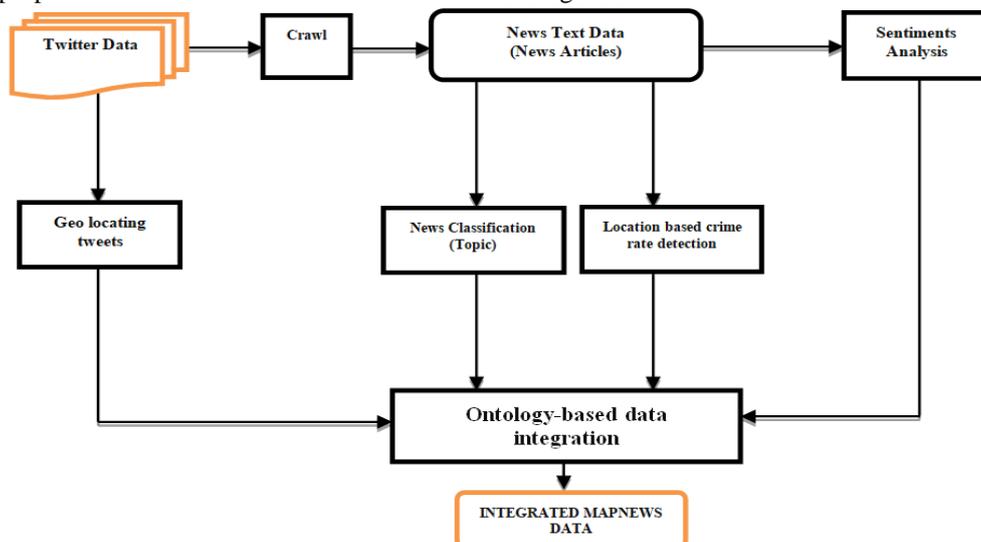


Figure 5.1: Framework for integrating map news data using ontology

The Proposed framework consists of five different modules as discussed below

A. Data-retrieving and location extraction

This module in turn contains two sub modules; in the first sub module live twitter data is being imported by manually providing the various news twitter handles, then link of each news article is extracted and using a web crawler entire article will be crawled and these articles which are unstructured real data are processed in structure news text data. Second sub module performed the Geo-locating of tweets obtained from twitter. Location extraction is performed in the form of latitude and longitude coordinates of each tweet using Geo-locating of tweets.

B. News article classification

The next module discusses the news classification of the structured news text data that is the outcome of the data - retrieving module. There are various steps involved in news classification such as parsing the articles, creating the features, training the model, getting the new data to feed to the model and showing useful insights. Here the model classifies the news articles into different categories like politics, entertainment, sports, tech and world news etc. For training the model, already classified datasets needs to be imported.

C. Location based crime rate detection using TF-IDF

This module uses the structured news text data of the data-retrieving module, detect the crime rate based on location, and display it in the form of a percentage. For crime rate detection in the news text data, a word dictionary related to crime is loaded and by using term frequency-inverse document frequency (tf-idf) each article checking will perform. Term frequency is the widely used technique in which for a particular term a weight is assigned based on total number of times that term is appearing in a news article or document. In this case, the imported crime word dictionary term checking is performing and weight is assigned based on its appearance in the news articles and location of that news article is extracted.

Sentiment Analysis of news articles

News articles may contain emotions like good, bad or neutral. Sentiment analysis is a way to examine the human emotions available in textual form. This module deals with the sentiment analysis of structured news text data obtained from the data-retrieving module. First, the news articles are preprocessed in order to perform tokenization. Tokenizing breaks news articles into individual words, symbols or phrases. After this tf-idf technique is used for term frequency calculation. This technique identifies the important words in a particular document, once the important words are identified, a dictionary is used for the identified words to assign to sentiment score based on their term frequency. The word dictionary used contains the number of different word senses that is used for sentiment score calculation. Each news article is treated as a document

and sentiment scores of each document is calculated. Based on the sentiment score each news article is labelled positive, negative or neutral. The article having scores of +1 is labelled as positive, the article having scores of -1 is labelled as negative and score 0 indicates the neutral article.

D. Ontology-based data integration

The last module deals with the integration of data, which are the outcomes of the above modules. Ontology is used for the purpose of integration of location and news articles originating from different sources. Ontology-based data integration is an approach used to combine data generating from multiple heterogeneous data sources. Data integration process may include the mapping of related data fields in order to resolve the variety challenge of news articles big data.

VI. CONCLUSION AND FUTURE WORK

Big data is at this point going all out to demonstrate its incentive in every field. In any case, to effectively use its potential in any field, there is a requirement for high handling tools to examine this various arrangement of complex information. Data creating through assorted gadgets comprises of various types and formats of data. To process and incorporate this heterogeneous information in numerous areas semantic technology gives some valuable and extendable help. Ontologies are the key piece of the semantic web through which mapping of various information fields can be performed. These advanced innovations are still at a beginning time and require further work in the fields where information incorporation assumes a significant job. The proposed framework helps in settling the difficulties in variety of big data utilizing semantic technology. Future work of this paper would be implementation of the proposed framework for the selected online news articles data which will integrate the heterogeneous data in order to resolve the variable challenge of big data.

REFERENCES

1. Qudamah Quboa, Nikolay Mehandjiev: "Creating Intelligent Business Systems by Utilizing Big Data and Semantics", 19th IEEE Conference on Business Informatics, Volume 2, 39-46, 2017.
2. K. Davis, D. Patterson, "Ethics of Big Data: Balancing Risk and Innovation", O'Reilly Media, 2012
3. vom Brocke, J., A. Simons, A. Clevn: Towards a business process-oriented approach to enterprise content management: The ECM-blueprinting framework. Inform. System-e-Bus, Manage, 9: 475-496. DOI: 10.1007/s10257-009-0124-6.
4. J Ahmed, R Gulmeher: "Nosql databases: New trend of databases, emerging reasons, classification and security issues", international journal of engineering sciences & research technology, Volume 4, Special issue 6, June 2015.
5. Berners-Lee, Tim, James Hendler, and Ora Lassila. "The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities," Scientific American, vol. 284, no. 5, pp. 1-5, 2001.
6. Red Hat Enterprise Linux: the five must-haves of big data storage, White Paper, 2013.

7. D. Calvanese, Martin Giese, Peter Haase, Ian Horrocks, T. Hubauer, Y. Ioannidis, Ernesto Jimenez-Ruiz, E. Kharlamov, H. Kllapi, J. Klüwer, Manolis Koubarakis, S. Lamparter, R. Möller, C. Neuenstadt, T. Nordtveit, Ö. Özçep, M. Rodriguez-Muro, M. Roshchin, F. Savo, Michael Schmidt, Ahmet Soylu, Arild Waaler, and Dmitry Zheleznyakov, "Optique: OBDA Solution for Big Data." In The Semantic Web: ESWC 2013 Satellite Events, pp. 293-295, 2013.
8. Ashley Vance, "Start-Up Goes After Big Data with Hadoop Helper," New York Times Blog, 2010.
9. Kouji Kozaki, "Ontology engineering for big data," in Proc. Of Ontology and Semantic Web for Big Data Workshop in the ICSEC 2013, September 4, 2013, Bangkok, Thailand.
10. Kharlamov, Evgeny, Ernesto Jimenez-Ruiz, Dmitry Zheleznyakov, Dimitris Bilidas, Martin Giese, Peter Haase, Ian Horrocks, Herald Kllapi, Manolis Koubarakis, Özgür Özçep, Mariano Rodríguez-Muro, Riccardo Rosati, Michael Schmidt, Rudolf Schlatte, Ahmet Soylu, and Arild Waaler. "Optique: Towards OBDA Systems for Industry," The Semantic Web: ESWC 2013 Satellite Events, pp. 125-140, 2013.
11. GoogleBigTable: <http://labs.google.com/papers/bigtable.html>
12. Richard Cyganiak and Anja Jentzsch. "Linking Open Data cloud diagram," <http://lod-cloud.net>, accessed on 4 April 2014.
13. Urbani, Jacopo. "On web-scale reasoning," PhD diss., Ph. D. Dissertation, Computer Sci. Dept., Vrije Universiteit, Amsterdam, Netherlands, 2013
14. Soylu, Ahmet, Martin Giese, Ernesto Jimenez-Ruiz, Evgeny Kharlamov, Dmitry Zheleznyakov, and Ian Horrocks. "OptiqueVQS: towards an ontology-based visual query system for big data," In Proc. of Proceedings of the Fifth International Conference on Management of Emergent Digital EcoSystems, ACM, pp. 119-126. 2013.
15. Hoppe, Anett, C. Nicolle, and A. Roxin. "Automatic ontology-based user profile learning from heterogeneous web resources in a big data context," in Proc. of Proceedings of the VLDB Endowment, vol. 6, no. 12, pp. 1428-1433, 2013.
16. Bizer, Christian, Peter Boncz, Michael L. Brodie, and Orri Erling. "The meaningful use of big data: four perspectives--four challenges," ACM SIGMOD Record, Vol. 40, no. 4, pp. 56-60, 2012
17. Margara, Alessandro, Jacopo Urbani, Frank van Harmelen, and Henri Bal. "Streaming the Web: Reasoning over Dynamic Data," Web Semantics: Science, Services and Agents on the World Wide Web, Vol. 25, pp. 24-44, 2014.
18. Nguyen, Trinh Hoang, Vimala Nunavath, and Andreas Prinz, "Big Data, Metadata Management in Smart Grids," In Big Data and Internet of Things: A Roadmap for Smart Environments, pp. 189-214, 2014.
19. Bennett, Mike, "The financial industry business ontology: Best practice for big data," Journal of Banking Regulation, vol. 14, no. 3, pp. 255-268, 2013.
20. Haase, Peter, Ian Horrocks, Dag Hovland, Thomas Hubauer, Ernesto Jimenez-Ruiz, Evgeny Kharlamov, Johan Klüwer, Christoph Pinkel et al. "Optique System: Towards Ontology and Mapping Management in OBDA Solutions," in Proc. of Second International Workshop on Debugging Ontologies and Ontology Mappings-WoDOOM13, p. 21. 2013.
21. Heath, Tom, and Christian Bizer "Linked data: Evolving the web into a global data space," synthesis lecture on the semantic web: theory and technology, Vol 1, no 1, pp. 1-136, 2011.
22. Shvaiko, Pavel, and Jérôme Euzenat, "Ontology matching: state of the art and future challenges," Knowledge and Data Engineering, IEEE Transactions on Vol. 25, no. 1, pp. 158-176, 2013
23. <http://www.oracle.com/technetwork/database/bigdata-appliance/overview/index.html>, accessed on 4 April 2014.
24. <http://hadoop.apache.org/>, accessed on 4 April 2014.
25. <http://www-01.ibm.com/software/data/bigdata/>, accessed on 4 April 2014.
26. <http://www.ontology.com/>, accessed on 6 April 2014.
27. Schätzle, Alexander, Martin Przyjaciół-Zablocki, and Georg Lausen. "PigSPARQL: Mapping sparql to pig latin," in Proc. of Proceedings of the International Workshop on Semantic Web Information Management, p. 4. ACM, 2011.
28. <http://sina-linkeddata.aksw.org/>, accessed on 5 April 2014.
29. <http://treo.deri.ie>, accessed on 3 April, 2014.
30. Whetzel, Patricia L., Natalya F. Noy, Nigam H. Shah, Paul R. Alexander, Csongor Nyulas, Tania Tudorache, and Mark A. Musen, "BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications," Nucleic acids research vol.39, no. Supply 2, pp. W541-W545, 2011.
31. Rysavy, Steven J., Dennis Bromley, and Valerie Daggett. "DIVE: A Graph-Based Visual-Analytics Framework for Big Data," Computer Graphics and Applications, IEEE, Vol. 34, no. 2, pp. 26-37, 2014
32. Szekeley, Pedro, Craig A. Knoblock, Fengyu Yang, Xuming Zhu, Eleanor E. Fink, Rachel Allen, and Georgina Goodlander. "Connecting the Smithsonian American Art Museum to the Linked Data Cloud," The Semantic Web: Semantics and Big Data, pp. 593-607, 2013.
33. Gottrath, Thomas, Ansgar Scherp, Bastian Krayner, and Arne Peters. "LODatio: A Schema-Based Retrieval System for Linked Open Data at Web-Scale," The Semantic Web: ESWC 2013 Satellite Events, pp. 142-146, 2013.
34. Konrath, Mathias, Thomas Gottrath, Steffen Staab, and Ansgar Scherp. "Schemex—efficient construction of a data catalogue by stream-based indexing of linked data," Web Semantics: Science, Services and Agents on the World Wide Web, Vol. 16, pp. 52-58, 2012.

AUTHORS PROFILE



Jeelani Ahmed pursued Bachelor of Engineering and Master of Technology in Computer Science from Visvesvaraya Technical University Belagavi, India in 2012 and 2015. He is currently pursuing Ph.D. From Maulana Azad National Urdu University, Hyderabad, India since 2017. His main research work focuses on Big Data Analytics, Semantic Web, Network Security and Cloud Security. He has 4 years of teaching experience and 2 years of Research Experience.



Dr. Muqem Ahmed is working as an Assistant Professor at the Department of Computer Science and Information Technology Hyderabad (India). He received his doctoral degree in Computer Science from Jamia Millia Islamia New Delhi India. His professional experience spans over more than 10 years of teaching, research, and project supervision. He has supervised various students in interdisciplinary research and industrial projects. Over the years, he has published many research papers to national and international journals of repute. In addition to these, he is also in the Editorial Boards and Reviewers' Panels of various journals. His primary area of research focuses on semantic web applications, Distributed Database Machine learning and big data Analytics.