

# A Critical Study of Semantic Analysis of Unstructured Web Data Using Big Data Technique

<sup>1</sup>Sachin Kumar Pandey & <sup>2</sup>Dr. Prabhat Pandey

<sup>1</sup>Research Scholar A.P.S University Rewa (M.P.) (India)

<sup>2</sup>Computer Science Department, A.P.S University Rewa (M.P.) (India)

---

## ARTICLE DETAILS

### Article History

Published Online: 12 June 2019

### Keywords

Big data, Process, Management.

---

## ABSTRACT

*Big Data, with their capability to give a profitable understanding into the improved basic leadership process, has as of late earned impressive profit by the two professionals and scholastic. Given the heterogeneity, the adaptability, the speed, the veracity, and the estimation of data, Big Data has developed where we face the real difficulties of procurement, accumulation, sifting, cleaning, investigation and perception of monstrous data. In this manner, these datasets must be abused at various dimensions with various systems. The development of Big Data is a primary test that necessities to break down an ever increasing number of individuals and understand a lot of data. So as to make the crude data utilized in basic leadership, suggestion frameworks, supposition investigation and area explicit use, it is prescribed to utilize Big Data Analysis to change and demonstrate the pertinent data, integrate and separate valuable concealed data. So as to have a superior outcome and to comprehend and identify connections among data and foresee future examples, the scientists brought the semantic viewpoint into the progression of Big Data Analysis. In this paper, we present a study that catches approaches which manage semantic for Big Data Analysis.*

---

## 1. Introduction

As of late, the amount of data has expanded exponentially. By 2020, it is hoping to have more than 16 zetta bytes valuable data Turner et al. (2014), where we are on the fringe of an age going up against the rise of another flood of data. The complete volume of data expended and offered on the web will increment by requests of loftiness Cavanillas et al. (2014). Consequently, the regular development of these datasets inside organizations requires new prerequisites as far as strategies, procedures, and devices identified with its handling and misuse Manyika et al. (2011). With this development of data, the Big Data idea has risen. At first, it is utilized to portray these enormous sets which by and large incorporate masses of unstructured data requiring high speeds to store, process and examine them. In this unique circumstance, we face the primary difficulties of obtaining, cleaning, incorporating, misusing, dissecting and imagining gigantic amount of data from broadly appropriated data sources. Along these lines, we accept that the undeniably gathered data is causing an issue in putting away and dealing with these tremendous and heterogeneous datasets. Likewise, it is important to process crude data to manage heterogeneity, adaptability, continuous, veracity and estimation of the Big Data. So as to make the crude data obtained usable in decisionmaking, expectation, proposal frameworks, supposition examination, it is prescribed to utilize the Big Data investigation. Truth be told, Big Data examination is a significant subdomain of Big Data Chen et al. (2014), without which the vast majority of the data acquired would be useless. It includes adding structures to the data as an initial step during or before data investigation. Thus, the more data are organized, the more effectively it will be handled by machines. The examination uncovered that the accompanying conventional procedures are either helpful today or in the short

and medium term: data mining, AI, data recovery, thinking, semantic handling, and data disclosure Domingue et al. (2016). By and by, there are new difficulties presented by the particular Big Data attributes: the volume, the assortment, and the speed that all the more explicitly include huge scale and high-speed thinking. Also the data can take diverse sentence structure organizations or mappings or various connotations connected to the equivalent syntactic structures. Semantic methods have show to be the most pertinent for tackling these issues. That is the reason specialists are centered around creating semantic-based Big Data investigation arrangements which can help give increasingly effective arrangements and better address the difficulties of Big Data examination for improved examination, basic leadership, forecast, data quality and separating new knowledge.

## 2. Definition of Big Data

The dimension of accord come to by an academic network on the meaning of an idea can be utilized as a proportion of the advancement of an order. As Big Data rather has advanced so quickly, there have been numerous endeavors to characterize Big Data, as far as use and citation Beyer and Laney (2012), Ward and Barker (2012), Microsoft (2013). In any case, none of these recommendations counteracted the creators of Big Data to overlook the past definitions and proposing new ones. Albeit Big Data is as yet a generally youthful idea, it absolutely merits an acknowledged vocabulary of references that permit the great improvement of the order. Among the current definitions, we will refer to a meaning of Gartner Bachlechner and Leimbach (2016): "Big Data is a term used to allude to high-volume, high-speed and high-assortment data resources that request practical, creative types of data handling for upgraded knowledge and basic leadership." The initial segment of the definition shows the three portion highlights of

Big Data imagined by Laney (these attributes are known as "3 Vs": the volume, the speed, and the assortment) Laney (2001). The second piece of Gartner's definition features the pertinence of both the connection among expenses and results and new mechanical abilities. At last, the third piece of the definition alludes to a definitive objective of making an

incentive through the preparing of data Lukoianova and Rubin (2013).

### 3. Big Data Challenges

In light of past looks into, it has been conceivable to recognize the difficulties related with Big Data. These difficulties are assembled into three classifications: provokes identified with data, procedures and the board.

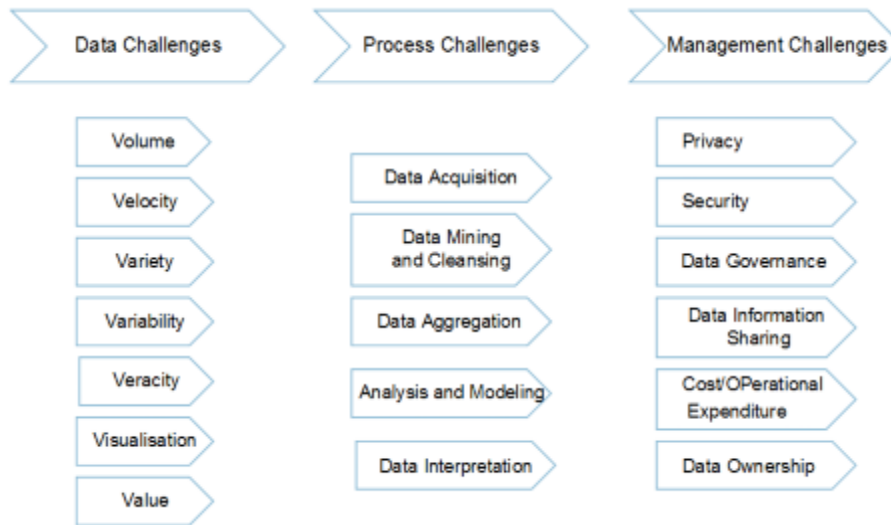


FIG.– Data Life Cycle

### 4. Data Challenges

The data difficulties are identified with the features of the data itself. Actually, the scientists have an alternate comprehension of the qualities of the data; some characterize just 3V [volume, speed, and variety] Shah et al. (2015), other speak to 4V [volume, speed, assortment, and variability] Liao et al. (2015), and some state they exist 6V [volume, assortment, speed, veracity, esteem and, variability] Sivarajah et al. (2017). At that point, the creators present the most significant qualities of the Big Data caught in the writing with an alternate rates: 39.64% for volume, 25.9% for assortment, 19,4% for veracity, 13.2% for worth, 7.9% for speed, 2.6% for perception lastly 1.8% for inconstancy. The seven qualities are depicted as pursues:

- **Volume:**The volume speaks to a huge test among the qualities of Big Data. It alludes to the measure of data that does not quit developing, and that is the perspective that strikes a chord of most people groups when they consider huge data containing sets of data including terabytes, petabytes of data or more. In 2012, an IBM study Schroeck et al. (2012) found that the greater part of the 1,144 respondents accepted that datasets of more than one terabyte were huge data. These datasets will achieve in excess of 40 zettabytes of data by 2020 Gantz and Reinsel (2012). For instance, Walmart creates around 2.5 terabytes of data in every hour of customer exchanges, and Facebook delivers more than 500 terabytes of data Zhao et al. (2013).
- **Variety:**The variety alludes to basic heterogeneity in a dataset. Truth be told, the associations utilize various sorts of organized, unstructured data and semi-organized. Organized data, which speak to just 5% of the genuine data Cukier (2010), alludes to

social databases or unthinkable data. The huge amount of data isn't cognizant and does not pursue a precise configuration or layout. It is caught in different structures containing unstructured data which speaks to 90% of all data Dobre and Xhafa (2014, for example, sound, video or messages on a few structures messages, tweets, and online journals or web data (for instance, pictures on informal communities, sensor data and considerably more Chen et al. (2012)). Thusly, XML is a solid case of semi-organized data which contains data marks characterized by the client who makes them lucid by the machine.

- **Veracity:**IBM created this component which speaks to the untrustworthiness that exists in many organized and unstructured data sources. It additionally suggests the absence of value and exactness, the vulnerability and the flighty idea of the data. Moreover, the veracity permits the clients understanding data in data sources. For instance, the sentiments of every client on various online networking systems are of questionable nature that suggests mediation the human Sivarajah et al. (2015).
- **Value:**Big Data specialists think about the value as a significant element. Be that as it may, data sources may contain important data and irrelevant data Zaslavsky et al. (2012). In this sense, it is a great idea to approach Big Data, yet it is important to change over them into value.
- **Velocity:**The velocity compare to the expanding speed at which data is made, handled, put away, dissected, or the speed at which new data are produced Gandomi and Haider (2015). Given the development of cell phones, the researchers need to deal with a huge number of consistent data sources

that require continuous investigation. In this specific circumstance, customary data the executives frameworks are not ready to deal with the enormous amount of data. So as to fill this hole, Big Data advancements come in. They enable organizations to oversee ongoing data from big volumes of data. For instance, Wal-Mart treats more than one million exchanges every hour Cukier (2010). These data can be included to deliver customized and continuous ideas for clients Sivarajah et al. (2017).

- **Visualization:**The data visualization is tied in with speaking to knowledge and data in a progressively neat and strong path by using diverse visual configurations, for example, illustrations or pictures Taheri et al. (2014). For instance, more than a great many items are sold on eBay every month. To make these data justifiable, eBay considered the Big Data visualization instrument, which can change mind boggling and huge datasets into portrayals. In light of these introductions, eBay's managers can envision the aftereffects of the inquiry by observing the most recent client criticism and examining the emotions Sivarajah et al. (2017).
- **Variability:**Among the seven component of Big Data, variability is another important trademark however frequently mistook for the variety. The significance of these data changes always and rapidly (for instance, Google or Facebook stores and produces numerous data groups). So as to lead a right notion examination, the specialists guarantee that the calculations must be able to comprehend and decode the precise feeling of a word in this setting Zhang et al. (2015a).

## 5. Process Challenges

For the most part, process difficulties depend on the presentation of the Big Data value chain. They speak to the progression of data in a Big Data association to fabricate valuable data from the data Cavanillas et al. (2014) Curry et al. (2014). The creators of Sivarajah et al. (2017) make an examination of 227 articles, to recognize the difficulties that establish the value chain from the data gathering stage up to the visualization of the outcomes with various rates. These difficulties are ordered by various values of rates: 42.7% for securing challenge, 16.7% for data mining and cleaning challenge, 12.8% for accumulation challenge, 11% for examination and displaying challenge, and 6.6% for the translation challenge.

- **Data acquisition:**Data acquisition is the method of gathering, cleaning and separating data before being put in a capacity arrangement in which the examination of the data might be activated. The acquisition of Big Data is frequently administered by four Vs which include high volume, high variety, high velocity, however low-value data Lyko et al. (2016).
- **Data Mining and Cleansing:**The second test is to concentrate and clean non-steady, missing, unstructured, and invalid data that can digress the aftereffects of the investigation. Normally, the structure of conventional data is characterized ahead of time and the data is pre-approved. In this way, the

need to deal with these conflicting data from outer sources has prompted the presentation of the data extraction step. In this progression, data have been changed into an organization that can be broke down by Big Data arrangement. In addition, a few setting approval standards can be utilized to clean and expel any invalid data.

- **Data Aggregation:**Data can be engendered over various datasets, which requires that data are gathered utilizing basic fields. This is the job of the collection and data portrayal arrange that prompts the mix of various data sets to accomplish a brought together perspective on clients, improving basic leadership and making new knowledge Erl et al. (2016).
- **Data Analysis and modeling:**The following test is Data investigation that comprises of investigating, demonstrating and changing data so as to feature the pertinent data, orchestrate and concentrate concealed data with an extraordinary potential Domingue et al. (2016). The most intriguing advance with regards to the value chain is the piece of the data examination that permits expelling pointless data Tilman et al. (2016). Big Data investigation is tied in with adding structure to data to help decisionmaking. Additionally, Big Data investigation makes crude data prepared for rebuilding, stockpiling, and use. In an ongoing exchange on data examination, Richard Benjamins in an individual correspondence stated: "Investigation without data is useless, investigation with awful data is perilous, examination with great data is the goal".
- **Data Interpretation:** The last test is the visualization and utilization of data. It covers datadriven business exercises that require achieving their investigation and the devices expected to incorporate data examination into their business. The ability to dissect gigantic amount of data and to discover helpful data wear little value if simply the experts that can decipher the outcomes. The Data Visualization stage worries to utilize instruments to graphically impart the examination results for a successful elucidation by business clients Tilman et al.(2016).

## 6. Management Challenges

The Big Data-related administration difficulties can be seen when overseeing data in touchy territories, for example, prescription, banks, protection and individual data. Associations must guarantee the security side for these workers and staff to envision just the pertinent data. In Sivarajah et al. (2017), five difficulties are exhibited; their relating rates are as per the following: 10.1% for protection, 7.5% for security, 4.4% for data and data sharing, 3.1% for cost/operational use lastly 1.3% for data possession. In this specific situation, the fundamental difficulties are:

**Privacy:** Data investigation can uncover private and classified data about people or associations. Along these lines, investigating such data can prompt security infringement.

**Security:** If the security issues are not taken care of, the Big Data won't have a significant thought. The security of Big Data includes guaranteeing that data systems are reasonably se-fix through approval and validation instruments. Big Data

security likewise includes the meaning of data access levels. Not at all like conventional social database the executives frameworks, NoSQL databases don't ordinarily give vigorous implicit security components Erl et al. (2016).

Data Governance: In data examination, associations must have Big Data the board and data administration structures Erl et al. (2016). Also, the nature of data concentrated on treatment by Big Data arrangements ought to be assessed. Consequently, creating low-quality data permits delivering low quality outcomes.

## 7. Semantics for Big Data

The data itself is gone up against with the Big Data issues, which presents difficulties in each progression of the value chain from the data gathering to the visualization and the utilization of that data. Consequently, we need a semantic setting to assist researchers with accessing to data and use and decipher the outcomes. In this unique situation, a similar term can be spoken to in various ways (for instance: assigning spots, articles or various individuals), and the outcome will rely upon their specific circumstance. On another side, we can discover various ideas speaking to a similar article. Then again, we can discover data which offer a specific definition that varies from another. Be that as it may, semantic innovations are utilized to determine the irregularities, the assessment and the revelation of new data from a current knowledge base. In this segment, we will address the various methodologies that consolidation Semantic with enormous data to make these data associated with this present reality.

### **Semantics for Big Data acquisition**

The semantic viewpoint can be added to the acquisition step that speaks to the way toward gathering, sifting and cleaning the data before putting them in a capacity arrangement. Also, the examination of the data can be activated to catch related and profitable data. It must be powerful and canny to catch helpful data and take out superfluous data that contains mistakes or irregularities. Consequently, successful logical calculations are expected to comprehend the wellspring of the data, process the data consistently and decrease the data before capacity Zheng et al. (2015). With this impact, we will make reference to certain methodologies that have featured these focuses. In Hussain et al. (2010), the creators proposed a plan to recover and store a lot of RDF data in Hadoop Distributed File System (HDFS). The framework procedures and puts RDF data in a record. At that point, the creators propose a calculation that decides the best inquiry plan expected to execute a few questions on various sizes of datasets. Along these lines, this framework is versatile and productive to effortlessly oversee billions of RDF significantly increases. In any case, it is tolerant of blunders and it doesn't give quick outcomes. The creators in Khodkh et al. (2010) looked at changed methodologies for handling RDF data. They proposed a robotized framework to process and examine inquiry about RDF data established in SPARQL inquiries, executing time, quantities of triples and diagrams contained in the info RDF document. They use different strategies that expansion the framework results, for example, the investigation of the info data. This framework comprises of three primary parts: the underlying part is data preprocessing, sought after by the inquiry improvement and handling part

which is done uniquely for basic inquiries. The last part speaks to the inquiry executing time investigation. In this part, the inquiries and their outcomes are put away in a database. Along these lines, if the client triggers similar questions on the equivalent dataset, the outcomes will be given legitimately. This capacity of the framework makes it like the activity of the reserve. By and large, the associations are unfit to move novel advances with a safe joining to their framework and data. Thus, they required to cutting edge Big Data advances with a develop usefulness that supports the interoperability, the reusability, and the movability. In Leida et al. (2016), the creators incorporate three layers of MDA (Model Driven Architectures): the CIM (computational ward model), the PIM (stage autonomous model) and the PSM (stage explicit model). These layers are then converted into a Big Data engineering that includes recognizing a lot of predefining (CIM). At that point, these models are mapped into the PIM and PSM (contingent upon the exact organization imperative). At long last, they can be kept running in the TOREADOR stage. In this work process, the framework upgrades the reuse, decreases improvement costs and gives a programming interface (utilizing the flash system that actualizes an issue tolerant appropriation in a bunch) to collaborate with dispersed memory. In any case, it experiences a contextualization issue that can be verified at the season of age; yet invalid when consolidating this announcement with different articulations which are produced in another unique situation.

### **Semantics for Big Data integration**

It identifies with the presentation of the semantic measurement after the data acquisition and purging advance in the coordination stage. It concerns the perfect mix and collection of unstructured Big Data, which normally need persuading data. These data coordination strategies endeavor to combine data from various sources and supply clients with a bound together perspective on the data that will be examined. In this subsection, we present methodologies of Semantics for Big Data joining. In Fang et al. (2016), the creators propose a methodology for enormous semantic data. It incorporates four layers: the metadata layer, which uses distinctive metadata standards, for example, MARC, DC, GILS to portray the travel industry data, for example, geographic area and connections between these assets. The cosmology layer permits guaranteeing semantic interoperability in various metadata types. To this end, we can utilize two strategies: the first permits incorporating the traits and ideas of the distinctive metadata rules into cosmology utilizing the OWL knowledge portrayal language. The second strategy permits utilize the ontological language to change the metadata position into RDF group. The Linked Data layer distributes the data as per the standard of Linked Data. It gives a brought together access instrument to various data configurations and produces semantic interoperability between these data. The data application layer permits giving the conventional watchword based recovery technique and more easy to understand interface for Interactive research. In Keller et al. (2016), the creators propose a framework for joining various wellsprings of air traffic the board data utilizing semantic combination procedures. It changes data from unique source groups into an institutionalized semantic representation in a triples store with sherlock data dependent on cosmology. In this unique

circumstance, SPARQL is utilized as an inquiry device to recover data about the inherent triples store. The fundamental design comprises of four stages: the initial step permits picking three starting points of the Sherlock for the joining: flight way data, airplane terminal climate data, and data reporting in real time Traffic Notices. The second step enables utilizing the ATM cosmology to include the semantic viewpoint. Tragically, the reality of requesting that aeronautical clients comprehend the ATM metaphysics and to gain proficiency with the SPARQL sentence structure isn't reasonable. The third step permits changing over the source organization of the first ATM data and making an interpretation of them into RDF significantly increases. The last advance permits questioning and downloading the administration; yet it has not yet been put executed. Issa et al. (2016) is another methodology that changes sensor data to semantic data utilizing the Semantic Sensor Network cosmology. The creators proposed in this examination a few strategies for stocking and preparing semantic data. The acquired semantic data are broke down to find valuable data. Moreover, Big Data innovations are utilized to change crude sensor data into semantic data and achieve the versatility required in escalated data situations. The utilization of the SSN (Semantic Sensor Network) philosophy attaches on semantic similarity to the sensor data. The principle issue here is the developing multifaceted nature of sensors conveyed in new frameworks.

### **Semantics for Big Data analysis**

Normally, enormous data sets contain generally unstructured data. In this way, huge scale handling of such semi-organized or unstructured datasets represents a critical test to the Big Data investigation Kaisler et al. (2013). In this subsection, we present the methodologies that incorporates semantic in the progression investigation. In Salakhutdinov and Hinton (2009), the creators portrayed a semantic hashing strategy that is utilized to discover paired codes. So as to fastly retrieve the reports, they utilize a Deep Autoencoder Network and the retropropagation calculation for finding semantically comparable records paying little heed to the archive measure. The creators of Huang et al. (2013) built up another model through the blend of the Latent Semantic Analysis (LSA) which is a factual strategy utilizing to find the covered up and fundamental (dormant) semantics of words in a corpus of reports by consolidating with a Deep Neural Network based structure to characterize web archives utilizing hashing procedures. In Liu et al. (2015), the creators displayed a multi-recolor Deep Neural Network strategy for a multi-space characterization, it utilizes a few procedures, for example, word hash, sacks of word and n-gram that frequently experience the ill effects of the issue of data scattering.. Zhang et al. (2015b) is another work that proposes an experimental examination on character-convolutive systems for the content grouping. The creators structured ConVnets utilizing an English thesaurus that was gotten from the WordNet where every equivalent word for a sentence or a word is grouped by semantic vicinity in the most usually comprehended sense Feinleib (2014). In Nural et al. (2015), the creators propose a methodology that permits picking a suitable model to investigate Big Data utilizing semantic advances. To this end, investigation metaphysics is created utilizing the Ontology Web Language (OWL) to help deduction for semi-computerized model choice. In the

examination stage, a various leveled work process is produced for the prescient investigation. The initial step of the work process chooses a viable model dependent on the properties of the datasets. At that point, the data are set up for examination in the second step. The third step refines the kind of models being utilized. For each progression, a lot of indicator factors are recommended. It guarantees preferences due to the way that they use metadata notwithstanding the properties processed from these data. Moreover, the creators incorporate space knowledge into the model determination issue for the investigation. The Scalation system decreases the expense of advancement time for a multi-worldview demonstrating task. In any case, it is important to have skill in ontologies and principle bases. In addition, the prescient examination regularly requires knowledge of structure of the experience that was utilized to produce the data and the decision of a displaying method often relies upon the reason for the investigation. In Banage et al. (2015), a cosmology based work process age approach is proposed to help to consequently create the work process. This methodology utilizes the Automatic Service Composition (ASC) to modernize the Cross-Industry Standard Process for Data Mining (CRISPDM) technique. Indeed, the ASC comprises of four fundamental advances: the arranging, the revelation, the choice and the execution. The creators center around the arranging step. They utilize the client needs and the properties of datasets to induce the work process for data investigation. In light of cosmology intended for CRISP-DM, two unique methodologies for the work process are created. Initial, a standard based methodology is proposed to include derivations. It actualizes Semantic Web Rule Language (SWRL) guidelines to find the theoretical administrations as per the properties of the datasets and the need of the client. Second, two use case situations are utilized in an inquiry based way to deal with demonstrate its practicability. Thus, more application-explicit unique work process can be produced utilizing this methodology. In Dayyani (2016), the creators utilize clever parts to configuration, create and execute a product stage. This stage can to a great extent dissect various budgetary data to help people in choices making. In this specific circumstance, seven layers are exhibited: the initial two layers permit gaining and separating the data. They compose the data as indicated by a semantic layer that incorporates three vital programming segments: the Technical Data Dictionary has practically all the philosophy and scientific categorization definitions, and different segments, the Fundamental Data Dictionary and the Data Conversion Agent. The fourth and the fifth layers dissect and model the data to build diverse examination sees. The last two layers control and present the examination perspectives to the client. The multi-layer engineering includes a comparable system origination established on the Web administration recommended by Nguyen et al. (2011) to make a versatile and adaptable multilayer programming stage for explicit spaces, for example, budgetary, coordinations and human services zones. This design gives convenient and programmed data from a crowd of data sources however they need steady examination to settle on speculation and business choices. Paik (2016) is another work that is identified with Big Data examination utilizing semantics. It proposes a system that gathers data from the web, remote gadgets or sensors. At that point, the structure dissects the gathered data and changes them into cases of the

space cosmology. It mechanizes the Cross-Industry Standard Processing for Data Mining (CRISP-DM) utilizing the Automatic Service Composition (ASC) to get ready and examine the data gathered in the initial step. A mechanization model of Big Data Analysis (BDA) with CRISP-DM utilizing ASC is utilized to show signs of improvement situational mindfulness. In the Perception layer, productive data is gotten utilizing BDA Intelligence. On the opposite side, the exhibition demonstrating is considered for any mindfulness in the bosses layers. In Yao et al. (2016), a novel system is created to join semantic techniques and the Big Data handling from security investigation in Big Data. Initially, the system gathers and preprocesses an enormous number of data from various data sources. At that point, it gives huge spaces to store the gathered data. From that point onward, a few devices and libraries are utilized to process the data. At long last, the structure depicts three kinds of investigation, data examination, for example, data mining, AI, and measurement examination strategies. The semantic investigation that depends on the utilization of cosmology and HCI examination which sends semantic examination and data investigation. This structure, it improves security investigation strategies, including continuous Big Data, computational execution, clump handling, data affiliation and data mining to meet the necessities of execution as far as data volume. So as to fill these holes, we propose a methodology that utilizes semantic memory to encourage

the semantic grouping of data including the normal for value and quality Taouli et al. (2018).

## 8. Conclusion

Big Data is a general idea for any gathering of gigantic, fast, and heterogeneous data sets. Consistently, moment and hour, the man and the machine rapidly produce data, regardless of whether we are not associated with our gadgets. Consequently, the enthusiasm of associations and organizations starts to value the significance of utilizing the Big Data more to help choices and game plans made to actualize their methodologies. Because of the Big Data, associations will be able to gather, reap, misuse every byte of appropriate data, and use it to take the better choices. Big Data advancements consider the answer for gigantic datasets, yet particularly the ability to comprehend and completely abuse its value. the best in class in semantic for Big Data examination and made some diagnostic and experimental correlations. This incorporated a general portrayal of every one of the methodologies just as a few disadvantages of them that can be treated in future inquires about. Truth be told, there are numerous new difficulties in huge data the executives, constant correspondence, expectation, and different difficulties and issues in overseeing, handling, and investigating Big Data.

## References

- Bachlechner, D. and T. Leimbach (2016). Big data challenges impact, potential responses and research needs. IEEE.
- Banage, T. G., S. Kumara, I. Paik, J. Zhang, T. H. A. S. Siriweera, and K. R. C. Koswatta (2015). Ontology-based workflow generation for intelligent big data analytics. IEEE.
- Beyer, M. and D. Laney (2012). The importance of big data: A definition. Gartner report.
- Cavanillas, Curry, and W. Wahlster (2014). In new horizons for a data-driven economy a roadmap for usage and exploitation of big data in europe. Springer.
- Chen, H., C. R., and S. V. C. (2012). Business intelligence and analytics: From big data. Volume 35, pp. 1105–1155.
- Chen, M., S. Mao, Y. Zhang, and V. C. Leungv (2014). Big data related technologies, challenges and future prospects. Springer.
- Cukier (2010). The economist data everywhere: A special report on managing information.
- Curry, E., A. Ngonga, J. Domingue, A. Freitas, M. Strohbach, and T. Becker (2014). Public deliverable of the eu-project big.
- Dayyani, B. (2016). Software architecture design and development of multi-layer highly modular platform using intelligent components for dynamic big data analytics. IEEE.
- Dobre, C. and S. F. Xhafa (2014). Intelligent sendees for big data science. Volume 37, pp. 267–281.
- Domingue, J., L. Nelia, F. Anna, S. M. Van Kasteren Tim, and T. Andreas (2016). Big data analysis. In New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe. Springer International Publishing.
- Erl, T., W. Khattak, and P. Buhler (2016). Big data fundamentals concepts, drivers and techniques. Prentice Hall Press.
- Fang, Y., Z. Jiaming, L. aohui, and G. Mei (2016). Semantic description and link construction of smart tourism linked databased on bd. IEEE.
- Feinleib, D. (2014). Big Data Bootcamp: What Managers Need to Know to Profit from the Big Data Revolution. Apress.
- Gandomi, A. and M. Haider (2015). Beyond the hype: Big data concepts, methods, and analytics. Volume 35, pp. 137–144.
- Gantz, J. and D. Reinsel (2012). The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. IDC EMC Corporation.
- Huang, P.-S., X. He, J. Gao, L. Deng, A. Acero, and L. Heck (2013). Learning deep structured semantic models for web search. CIKM '13, New York, NY, USA, pp. 2333–2338. 22nd ACM international conference on information; knowledge management.
- Hussain, M., L. Khany, M. Kantarcioglu, and B. Thuraisingham (2010). Data intensive query processing for large rdf graphs using cloud computing tools. IEEE.
- Issa, H., L. van Elst, and A. Dengel (2016). Using smartphones for prototyping semantic sensor analysis systems.
- Kaisler, S., F. Armour, J. Espinosa, and W. Money (2013). Big data: Issues and challenges moving forward. pp. 995–1004. 46th Hawaii International Conference on System Sciences (HICSS).
- Keller, R., S. Ranjan, M. Wei, and M. Eshow (2016). Semantic representation and scale-up of integrated air traffic management data. In Proceedings of the International Workshop on Semantic BD, ACM.
- Khodkh, P., S. Lawange, A. Bhagat, K. Dongre, and C. Ingole (2010). Query processing over large rdf using sparql in big data. Communications of the ACM.

23. Laney, D. (2001). 3d data management: Controlling data volume, velocity, and variety.
24. Leida, M., C. Ruiz, and P. Ceravolo (2016). Facing big data variety in a model driven approach. IEEE.
25. Liao, Z., Q. Yin, Y. Huang, and L. Sheng (2015). Management and application of mobile big data. Volume 7, pp. 63 – 70. International Journal of Embedded Systems.
26. Liu, X., J. Gao, X. He, L. Deng, K. Duh, and Y.-Y. Wang (2015). Representation learning using multi-task deep neural networks for semantic classification and information retrieval. NAACL.
27. Lukoianova, T. and V. Rubin (2013). "veracity roadmap: Is big data objective, truthful and credible?". Volume 24, pp. 4–15.
28. Lyko, K., M. Nitzschke, N. Ngonga, and C. Axel (2016). Big data acquisition. In A Roadmap for Usage and Exploitation of Big Data in Europe. Springer International Publishing.
29. Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. Byers (2011). Big data: The next frontier for innovation, competition, and productivity. pp. 156.
30. Microsoft (2013). The big bang: How the big data explosion is changing the world.
31. Nguyen, T., A. Colman, and J. Han (2011). Modeling and managing variability in processbased service compositions. Service-Oriented Computing. Nural, M., M. Cotterell, and J. Miller (2015). Using semantics i Paik, I. (2016). Situation awareness based on big data analysis. International Conference on Machine Learning and Cybernetics (ICMLC).
32. Salakhutdinov, R. and G. Hinton (2009). Semantic hashing. International Journal of Approximate Reasoning 50(7), 969 – 978. Special Section on Graphical Models and Information Retrieval.
33. Schroeck, M., R. Shockley, J. Smart, D. Romero Morales, and P. Tufano (2012). Analytics: The real-world use of big data. how innovative enterprises extract value from uncertain data. IBM Institute for Business Value.
34. Shah, T., F. Rabhi, and P. Ray (2015). Investigating an ontology-based approach for big data analysis of inter-dependent medical and oral health conditions. Volume 76, pp. 351 – 367. Cluster Computing.
35. Sivarajah, U., Z. Irani, and V. Weerakkody (2015). Evaluating the use and impact of web 2.0 technologies. Volume 32, pp. 473–487.
36. Sivarajah, U., M. M. Kamal, Z. Irani, and V. Weerakkody (2017). Critical analysis of big data challenges and analytical methods. Volume 70. In Journal of Business Research.
37. Taheri, J., A. Zomava, H. Siegel, and Z. Tari (2014). Pareto frontier for job execution and data transfer tune in hybrid clouds. Volume 37, pp. 321–334.
38. Taouli, A., D. AmarBensaber, N. Keskes, and K. Bencherif (2018). Semantic classification for big data analysis. 14th International Conference on Business Intelligence and Big Data,