13.0 Predictive Analytics for Network Big Data using Knowledge Based Reasoning for Smart Retrieval of Data, Information, Knowledge, and Wisdom (DIKW)

Aziyati Yusoff
Universiti Tenaga Nasional, Malaysia
aziyati@mohr.gov.my

Norashidah Md Din
Universiti Tenaga Nasional, Malaysia
norashidah@uniten.edu.my

Salman Yussof
Universiti Tenaga Nasional, Malaysia
salman@uniten.edu.my

Assad Abbas
North Dakota State University, USA
assad.abbas@ndsu.edu

Samee U. Khan
North Dakota State University, USA
samee.khan@ndsu.edu

Abstract

Data, Information, Knowledge, and Wisdom (DIKW) Hierarchy is represented by a taxonomical pyramid illustrating its significance and the key role played by the sets of data. The pyramid becomes complex when the big data floods out of the storage stream and challenges the wisdom of the hierarchy. The reason is the incompetence of the systems deployed across the networks in managing large but heterogeneous data volumes. Therefore, besides deployment of efficient computing methodologies to deal with the growing data, defining the strong relationships to infer the meaningful information by applying the wisdom is of significant importance. In this chapter, the wisdom from the pyramid which is obtained from the root knowledge and information within the online services of big data is elucidated by using the knowledge based reasoning. This generalization approach is further elaborated through statistical inferences and analytics including the process of making hypotheses, assumptions, and normality of distribution study, especially for the numerical data. Moreover, the chapter demonstrates the feasibility of designing the ontological arguments to support the reasoning behind the DIKW hierarchy. The ontological arguments help in validating the relationships and propositional logic behind the big data. The use of big data, for example in the area of disaster management in Malaysia is exemplified thereby proving the hypothetical statements and the use of statistical predictive model.

13.1 Introduction

The study of Data Science, Analysis, and Decision Making has evolved from translating the raw data, information sharing, and knowledge representation to the wisdom of Web of Things. Starting from the idea of architecting a wisdom hierarchy, the base of the hierarchy is built upon data, information, knowledge, and wisdom pyramid [1]. The pyramid or hierarchy as illustrated in Fig. 1 consists of the components of Data, Information, Knowledge and Wisdom (DIKW). In addition, the recent trend on the needs of network big data has challenged this hierarchy to be redefined and implemented beyond the contemporary use of data analytics. If data on its own is raw; information is adding the context; knowledge is describing on how to use it; and wisdom is explaining on why to use it, [2] then the big data is challenging the hierarchy to be in a more complex yet integrated structure.

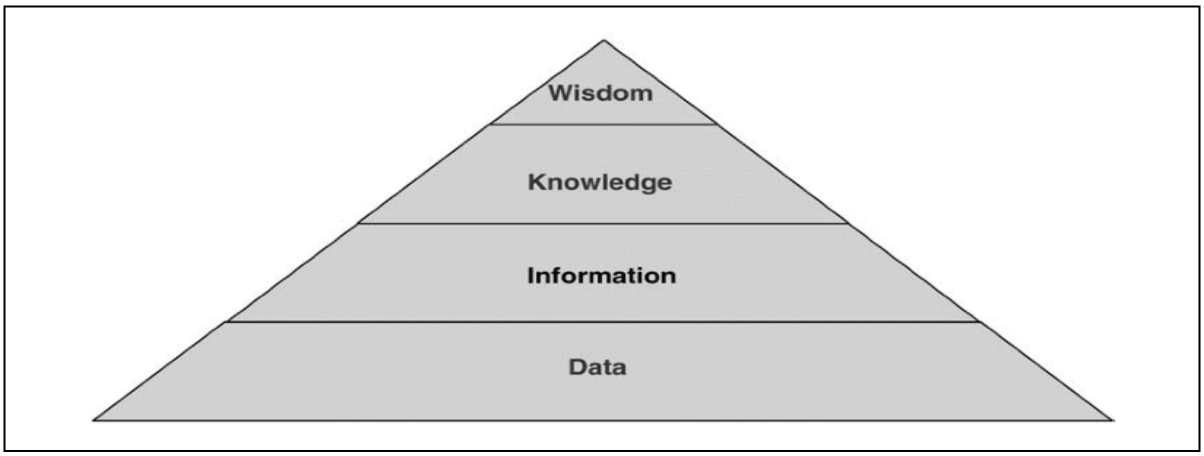

Fig. 1. The DIKW hierarchy

In this chapter, the first section describes the background of DIKW and the challenges that it poses to the big data network. The second section is deliberating on the form that most of big data being represented in enterprises such as the mathematical model, statistical inferences, machine learning and decision making techniques. Hypothesis testing was included to illustrate the engineering process of this statistical method. Section 3 discusses the knowledge based reasoning that can be induced from the statistical method that was performed in the previous section. The integration of these two scopes of study is constructed by the technique of ontology and semantic network. Therefore, from this technique the predicate calculus, propositional logics and knowledge representations are highly enumerated. The platform of ontology that is used through out this study is by using the Web Ontology Language (OWL). The OWL is used to design the big data feed and expected to demonstrate a quality performance for the use in the platform of DIKW. Section 4 of this research demonstrates on the smart retrieval that can be benefitted from the OWL design and performance previously described on preceding section and Section 5 concludes the chapter.

13.2 Data, Information, Knowledge and Wisdom (DIKW) Hierarchy in Network Big Data

13.2.1 What is DIKW?

The term DIKW was introduced in 1980s [3, 4] where originally, the wisdom phase was not proven because it was inferred that wisdom would deal with values and judgments. The ability of dealing with values and judgments was not foreseen in computer and automated machineries at that time [5, 6]. Data is an important element in the built of a computer [7]. From data, people found ways provide further information on subject matter. For the past 15 years, there has been tremendous progress in computational knowledge management [8]. However, the mathematical and computational sciences never stop at knowledge sharing methodologies. Five years later, the wisdom of DIKW hierarchy came into limelight [9, 10].

13.2.2 DIKW and the Challenges in Big Data Network

As the big data is making its wave in 2010s, people are making a revisit to DIKW. When the data was treated as the subject of matter for all of computing machineries, they were organized and architected such that it can be represented in a manageable way for storage and retrieval purposes. However, with the inception of cloud computing as an alternative storage, the types of data have evolved from structured to semi-structured and unstructured forms. This type of data is now known as the big data.

The term big data is always associated with its characteristics generally called as 5Vs. The 5Vs include Volume, Velocity, Variety, Veracity, and Value [11, 12]. The data science was originally considered mostly as technical but as the enterprises demanded more in terms of data storage and processing power, big data enabled methods started emerging rapidly. Therefore, the big data operator should have the ability to accommodate the variety of data structures, the volume it carries, the velocity for information retrieval, the veracity of its users' behavior, and the value that it represents.

In addition, if the big data is to be operated for the purpose of information sharing and knowledge management, then the DIKW hierarchy is best to be considered for the system architecture and smart retrieval engine. The aforementioned operational study is also known as network big data. The analytics among other things involve the intelligent phase, the design phase, and, choice phase. As a performance measure, the implementation of analytics depends on the success of the engine to validate and verify the problem for decision making process.

13.2.3 The Network Big Data Framework and Architecture

The network big data in this chapter will be discussed in two case studies. Case study I is about flood information management in the state of Kelantan, Malaysia and Case Study II is about the demand and skills qualification of labor market in Malaysia. The method of analysis for these two case studies will involve the statistical inferences and knowledge based reasoning in network big data prediction.

Fig. 2 is illustrating the general framework of network big data for DIKW smart retrieval.

The framework consists of the input parameters, the process involved and the output parameters. The input parameters are also known as the design parameters are all the possible big data feed. The process involved is also referred to as the engine consists of statistical inference approach and knowledge based reasoning methodology. While the output parameters are the performance parameters and are the objectives of this study i.e. the DIKW Smart Retrieval.

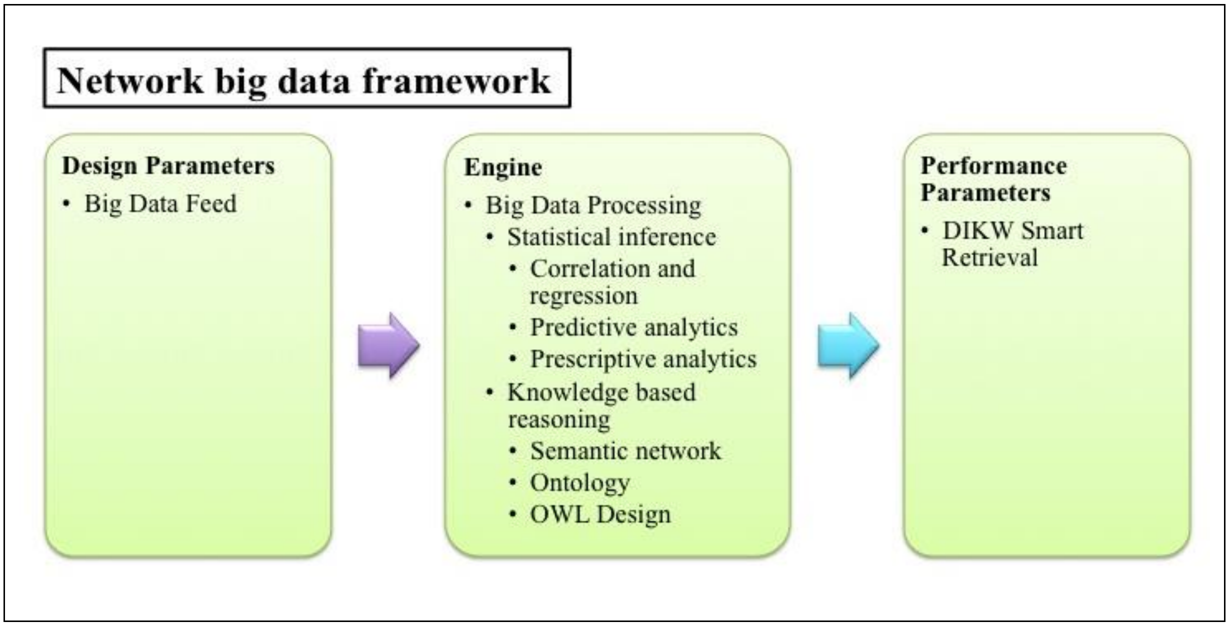

Fig.2. The Network Big Data Framework for DIKW Smart Retrieval

The challenge in network big data is mainly about how the data is read, processed and translated. The big data comprises all kinds of data types and high in complexity. This complexity is also challenged by the anticipation of big data within the Internet technology. As a result, the environment of network big data is dynamic and should be able to transform the data feed into the desired form of data presentation. In this chapter, the method of handling network big data is deliberated by the approach of Smart Retrieval engine. This engine is aimed to sort the data in accordance to the DIKW hierarchy.

In addition, the architecture for this network big data demonstrated in both of the case studies is as illustrated in Fig. 3. The network big data architecture is built upon three main parameters, i.e. the methodology, design and performance.

The methodology component comprises of all forms of data readings including the types of *.txt* for text data, *.csv* for numerical data, *.kml* for geographic-based location data, and *.script* for online data services. The design component is built upon all possible elements of a computing program including the algorithm, tools and interface. The algorithm involved is connotation to ontology engineering design consists of propositional logic and predicate calculus. The tool that used for this study is the Web Ontology Language (OWL). The interface for this design is the Uniform Resource Identifiers (URI) and the Resource Description Framework (RDF). As for the performance of this study, the architecture is explicitly showing that the two main elements are the smart retrieval by using the hierarchical factor of DIKW.

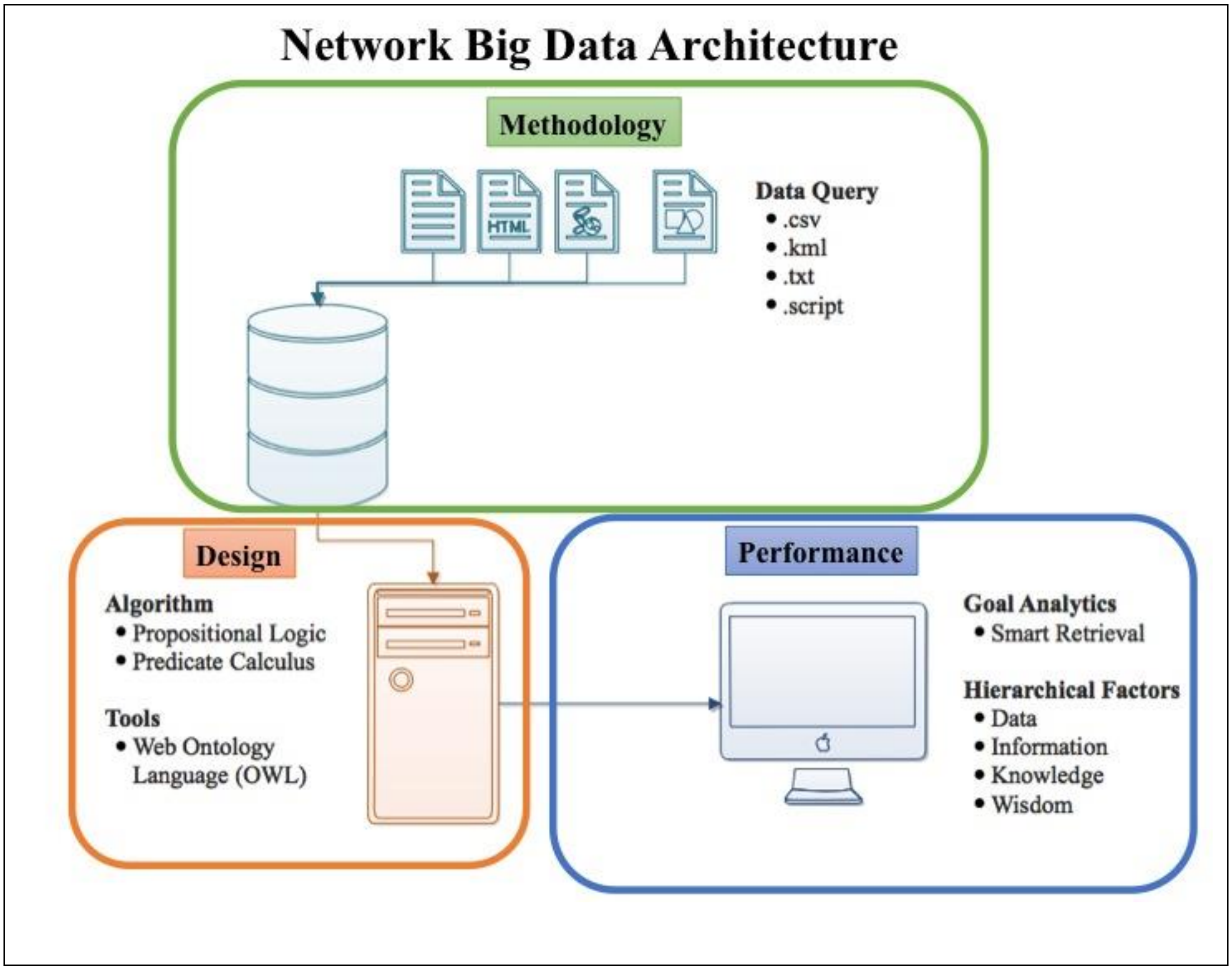

Fig. 3. The Network Big Data Architecture for DIKW Smart Retrieval

13.3 Statistical Inferences and Analytics in Network Big Data

13.3.1 Hypothesis Testing for Big Data

One of the methods that usually entrepreneurs, data scientists, and engineers apply for decision making process is statistical analysis. When a study involves either of the nature of relationship to a certain subject matter, the differences among groups, or the independence of variability between two or more factors, the approach taken is usually by hypothesis testing [13]. In the light of above big data methods seem suitable to be evaluated for their relationship from one component to another, or from one class to another, or from one category to another. This is further proven by the needs of the analytics on data collection, analysis and, its interpretation.

13.3.2 Correlation and Regression Analysis

To illustrate the application of hypothesis testing, two case studies were chosen. The hypotheses will concentrate on correlation and regression analysis. The correlation and regression analysis is done to study on the relationships and the strength of the connections between the involved parameters. Therefore, it fits the study of data analysis towards the hierarchy in interpreting the information, knowledge, and wisdom of the findings. Case study I is about the findings in flood management in the State of Kelantan, Malaysia and Case Study II is about the findings in the demand and skills qualification of labor market in Malaysia.

(i) *Case Study I: Flood management in the state of Kelantan, Malaysia*

A case study was carried out to the flood disaster in the state of Kelantan, Malaysia. This disaster occurrence was seen to be an annual event that many researchers keep coming back to the state to study on its causes and implications. The flood is always happening during the North-Eastern Monsoon which usually takes place between October to February every year. One of the studies had suggested that the flood keeps happening due to the volume of the rainfall at that specific time of the year which had caused the state to be in constant pour for several days and contributed to the rise of water level. Hypothesis statements were made on this fact including the alternatives. However, the end result was quite surprising that the flood was not totally relying upon the pour of the rainfalls. The relationship measured for this hypothesis was showing a weak connection. This has made the researchers to further investigate on other factors that might trigger the cause of the flood [14].

The sample readings of correlation and regression for this case study were illustrating the results during the flood period in the state of Kelantan [15]. However, the graphical representation of water level with respect to rainfall is showing some scattered data from the correlation readings. This did not represent the strong relationship between the two variables to induce them as the cause of the flood incident.

(ii) *Case Study II: The demand and skills qualification of labor market in Malaysia*

Another case study was done on best practices and comparative analysis on the labor market in Malaysia and the training institutions that are responsible to fulfill her industry needs and demand. Like most of developed and developing countries, Malaysia is also coping with the needs of highly skilled manpower resources to work at the plants and industries. To administer the quality skilled workforce, Malaysia has its own government agency that manages the Occupational and Skills Qualifications. The main aim of this agency is to certify the skilled workers with the right level of certification.

The hypothesis made for this study was on the relationship of a good quality of skilled workforce and its dependency to the Occupational Analysis (OA) and the curricula that were founded and developed by the governing agency. The study was carried out by interview sessions with focus groups. As a result, most of the respondents agreed that to produce knowledge workers, the OA and strong curricula indeed have a high influence on the successful implementation of training institutions and should be able to fulfill the needs of the industry [15]. This result is contrarily shown as compared to the Case Study I because the regression value of the result computation is showing a strong relationship between the two variables.

The above two case studies demonstrated the use of correlation and regression to a hypothesis. This approach is usually dealing with numerous aspects of relationship in social sciences scope of studies. Though the parameters might be technical and scientific, this correlation and regression is used to prove on the relationship in society's problem solving. However, correlation is not causation [16]. The study on correlation and regression does not provide insight to the analyzed data and information. Knowledge might be extracted but limited to some extend. Therefore, the hypothesis testing, correlation, and regression seem feasible at the foundation level of a wisdom abstraction. In consequence, further investigations, techniques and computational approach are needed to process the wisdom of subject of matter.

13.3.3 Predictive Analytics

Before a computer was created, perhaps the term 'prediction' was more likely to be used by the astrologers and clairvoyant [17]. It was considered a taboo to some cultures and a religious ritual to some other. In statistical analysis, data prediction can be performed by identifying the variables, parameters, and environments. Graphical representations usually involve the curve of normal distribution. From the normal curve, the study on its properties and characteristics may lead the researcher to predict on the next data reading. This prediction method is only carried out by comparing the pattern of previous data sets of the subject matter.

To illustrate on the predictive analytics, Case Study I is referred. The study on predicting the flood in the state of Kelantan had long been done due to the nature of its occurrence that takes place almost annually. The state of Kelantan, Malaysia has the area of 117 km squared of land area bounded by the latitude from 6^{O}7'N to 6^{O}14'N, and the longitude from 102^{O}14'E to longitude. Kelantan River is the largest river of the state and has a delta built but poorly drained [18].

Predicting the flood disaster from statistical hypothesis is similar to forecasting computation to a normal distribution curve. However, the challenge to the data readings to this case is the aim of the study is to reduce the disaster to zero. This is unlike the normal forecasting process that is used to maximize profits to specific business purposes.

In predicting the flood incident, several assumptions and limitations are made to define the scope of the study. The assumptions to this case study include: (i) the rainfall data readings are of normal distribution, (ii) the water level has very high dependency on the volume of rainfall, and (iii) other relative independent variables are very small and negligible. In addition, the limitations to the research parameters should also consider other factors for the occurrence of this incident including the wind movement, monsoon, global phenomena, tidal wave, and the gravitational force of moon and earth that affects the rise and fall of sea level.

The above-mentioned assumptions and limitations to the case study are best presented in similar Geographical Information System (GIS). Simultaneously, the result of the prediction will be illustrated in images format, such as geospatial mapping or Keyhole Markup Language

(KML). The simulation of the rise of water level is also expected to run if the calculation of the prediction is accurate by using the accurate parameters of data prediction. On the other hand, such kind of application is seen as static and provide sole data to the disaster alert but contains less information, knowledge, and wisdom that would be able to aid not only the authorizing agencies but also the affect communities as a whole.

In early 2000s Google Inc. started to use an engine known as Hadoop MapReduce [19]. Hadoop is an open-source Java based programming language that aids to process massive data sets in a distributed environment [20]. This engine is also responsible to predict the search term that is expected from the user input. As a comparison, the approaches of predicting the output by using Hadoop MapReduce in Java platform and statistical analysis differ in many aspects yet able to be interconnected. Hence, this study is deliberating on the keen of integrating the source of statistical computation that can be applied to a machine learning platform such as an ontology. This will be further discussed in the next sections.

13.4 Knowledge Based Reasoning of Data Prediction in Network Big Data

13.4.1 Semantic Network and Ontology of Big Data

A semantic network is used to show the connections from one knowledge to another as a set of concept. It was also used as a graph structure to represent knowledge in patterns of interconnected nodes and arcs which were first developed for artificial intelligence and machine translation [21, 22]. The semantic network is favorably used for knowledge representation in ontologies. An ontology is a way to formally model the structure of a system. This includes the definitions of the relationships and concepts within the system and usually the representation of logic in the form of propositional logic and predicate calculus [23, 24].

As an application to the business websites and online enterprises, semantic web can be created by using a formal ontology approach. This semantic web is able to extract, analyze, and manipulate the data in accordance with the business requirements [25]. One of the languages that is used widely for this purpose is the Web Ontology Language (OWL). The OWL is a semantic markup language for publishing and sharing ontologies on the World Wide Web [26] and the common syntax include the Extensible Markup Language (XML) or Resource Description Framework (RDF).

13.4.2 OWL Design for Online Prediction

To illustrate the OWL design for online prediction, the Case Study I is revisited. Defining its semantic network performs the OWL design for this case study. Classes and subclasses are identified and the relationships between one class to another are defined. This is as illustrated in [27]. The semantic network of flood management is concentrated from the study of the big data it carries. The big data of this case is annotated by main class and sub-classes including Hydrology Data, Tidal Wave and Monsoon Data, Telemetry Data and Meteorology Data.

From the classification modeling above, the properties of each of subclass with its main class and instances can be defined thereby illustrating the relationships of the model. The relationships of the above examples include object properties, sub-properties, and functional properties such as:

(a) *hasData* (NetworkBigData, (HydrologyData, TelemetryData, TidalWaveMonsoonData, MeteorologyData))
(b) *hasData* (HydrologyData, (Rainfall, WaterLevel))
(c) *providesAlert* (HydrologyData, WaterLevel)) & *hasData* (WaterLevel, (GunungGagau, RantauPanjang, Jeli, Dabong, Laloh, GuaMusang, KualaKrai, Kusial, Tualang, Aring, Jenob, PasirPutih, KotaBharu))
(d) *hasInformation* (NetworkBigData, HydrologyData) & *hasInformation* (HydrologyData, WaterLevel) & *hasInformation* (WaterLevel, (GunungGagau, RantauPanjang, Jeli, Dabong, Laloh, GuaMusang, KualaKrai, Kusial, Tualang, Aring, Jenob, PasirPutih, KotaBharu))
(e) *mitigateAction* (FloodInformationManagement, NetworkBigData) & *emergencyResponse* (FloodInfomationManagement, (WaterLevel, (GunungGagau, RantauPanjang, Jeli, Dabong, Laloh, GuaMusang, KualaKrai, Kusial, Tualang, Aring, Jenob, PasirPutih, KotaBharu)))

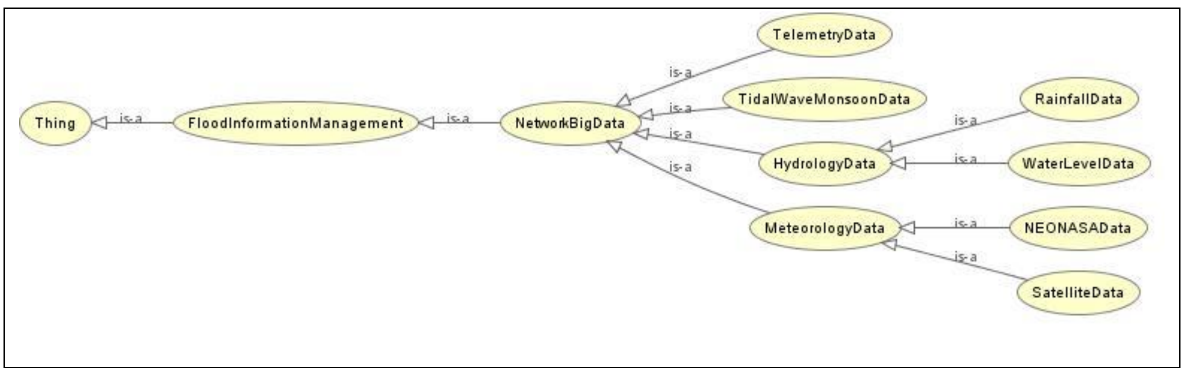

Fig. 4. The ontology for NetworkBigData for FloodInformationManagement

Consequently, an ontology architecture of the above mentioned semantic network is inferred as illustrated in Fig. 4. The ontology is showing the relationships of the instances with the subclasses and the main class of the study. This ontology relationship is further translated to the predicate calculus. The predicate calculus for the above illustrated ontology among others include:

(a) $\forall x1$: (HydrologyData ($x1$) => NetworkBigData ($x1$))
(b) $\forall x2$: (TelemetryData ($x2$) => NetworkBigData ($x2$))
(c) $\exists x3$: (MeteorologyData ($x3$) => NEONASAData ($x3$))
(d) $\exists x4$: (WaterLevelData ($x4$) => HydrologyData ($x4$))
(e) $\forall x$ P (x) $\Leftrightarrow$ P ($x1$) $\wedge$ P ($x2$) $\wedge$ P ($x3$) $\wedge$ P ($x4$)

Hence, it is from this process of designing the semantic network, the identification of properties, sub-properties, and instances, the design of ontology relationships, and the translation of predicate calculus that the engine of smart retrieval is instructed and operated. In general, the translation of the above ontology engineering study is using the method of Web Ontology Language (OWL).

13.4.3 OWL Performance for DIKW and Beyond

The ultimate performance for an OWL design is to produce a semantic Web application using the XML or RDF syntax and data interchange. The advantages of using the OWL design for semantic network architecture include the ability to identify the conceptual modeling, the relationship representations, and the object properties.

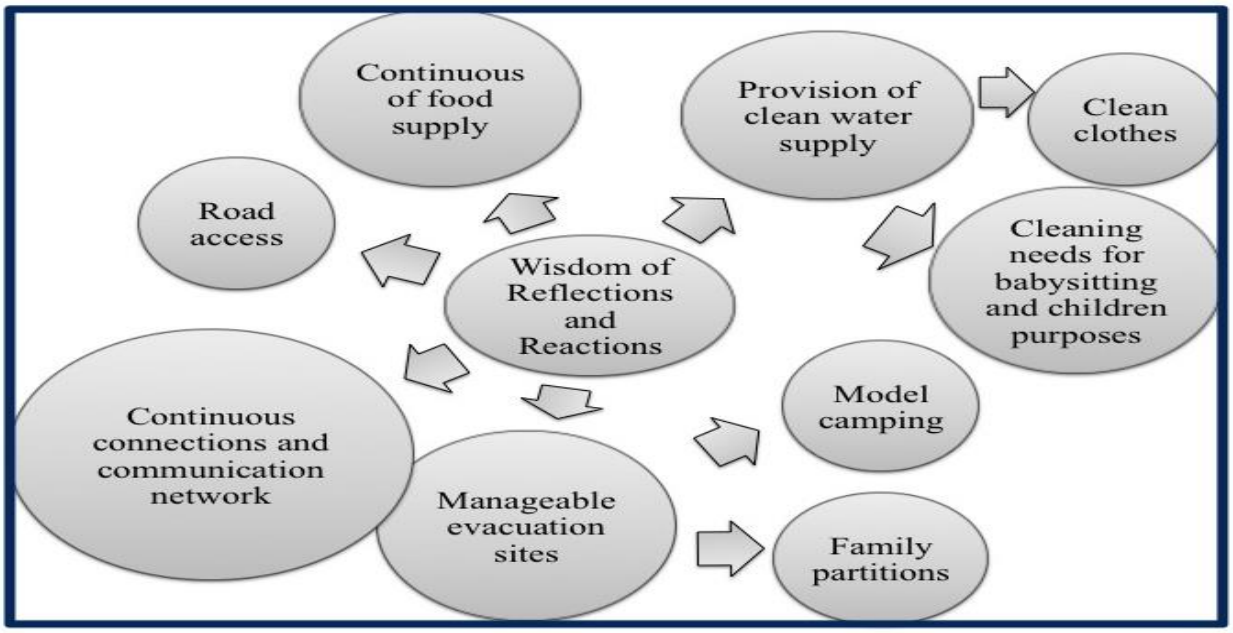

Fig. 5. The Wisdom of Reflections and Reactions to the Flood Incident

As for the Case Study I, the performance analysis is about the actions that are able to be implemented once the flood prediction is activated. This is as illustrated in Fig. 5. The main aim of the study is to generate the wisdom that can be abstracted from the data, information, and knowledge of the flood readings at the respective reading stations. Consequently, the wisdom in action when this disaster happen will include: (i) provision of clean water supply, (ii) manageable evacuation sites, (iii) continuous connections and communication network, (iv) road access, and (v) continuous of food supply.

As a summary, for both of the previous case studies, the OWL performance is expected to provide services which should comply with the initial objectives of the study. The design, performance, and engine for the case studies are illustrated in Table 1. The OWL performance is analyzed in the form of DIKW perspectives. This summary is also to illustrate that the integration of prediction analysis with the DIKW hierarchy can be performed successfully.

Table 1: The generic details of big data, design, engine and performance parameters of Case Study I and Case Study II.

Case Study	Case Study I: *Flood management in the state of Kelantan, Malaysia*.	Case Study II: *The demand and skills qualification of labor market in Malaysia.*
Big data	Hydrology Data (rainfall and water level), GIS Data (geospatial mapping, location), Meteorology Data (longitude latitude, monsoon).	Technical and Vocational Education Training (TVET) Data in selected countries i.e. Malaysia, Singapore, Australia, and Canada.
Design	OWL Design which include the following parameters: (i) rainfall data (ii) water level data (iii) telemetry data (iv) NEO NASA data (v) tidal wave and monsoon data	OWL Design which include the following parameters: (i) occupational framework (ii) occupational competency standard (iii) TVET qualification (iv) TVET Standards & Curriculum Guide (v) TVET term used (vi) industry partnership (vii) training institutions
Engine	Smart web using the taxonomy of disaster management in XML syntax and Uniform Resource Identifier (URI).	Smart web based on the triangulation method using the design parameters to integrate the industry demands, training needs analysis and the number of skilled workforce that are able to be produced.
Performance	OWL Performance which include the following parameters: (i) *Data*- water level readings (ii) *Information*- alert on the flood to occur within hours or days. (iii) *Knowledge*- mitigation actions, emergency response and rescue from the authorities, access road, and evacuation centers. (iv) *Wisdom*- the number of nearby evacuation centers with details of its distance, capacity, etc., the nearest access road, available transport, and the duration of moving.	OWL Performance which include the following parameters: (i) *Data*- number of training institutions and occupational standards developed. (ii) *Information*- best practices of TVET by comparative analysis in selected countries, and recommendations in standard methodology. (iii) *Knowledge*- TVET implementation and curriculum developed that are able to fulfill the needs of labor and industry demand. (iv) *Wisdom*- Standard methodology for occupational frameworks, and competency standards.

13.5 Smart Retrieval Prediction Engine in Network Big Data

In comparison to the study of Information Retrieval (IR), the issues that are affecting the performance of this area of study include the ability of crawlers, indexing, and ranking [28]. Designing a smart retrieval online prediction engine in ontology platform is seen to be the next frontier to the available IR engine. The constraint of this application is to retrieve the alert of flood incident in the form of DIKW hierarchy.

This is as illustrated in Fig. 6. The network big data for this system involves the process of sentiment analysis for social media network, statistical inferences, and web crawling from the available authorities web applications. The analyzed data will undergo the architecture of semantic network, the OWL design, and identification of propositional logic in knowledge representation. These three processes are the fundamental structures to the online prediction engine. The syntax of XML or RDF from the OWL design is strengthening the operations of this engine. As an output, the engine is expected to be able to process the network big data to produce the DIKW of flood incident.

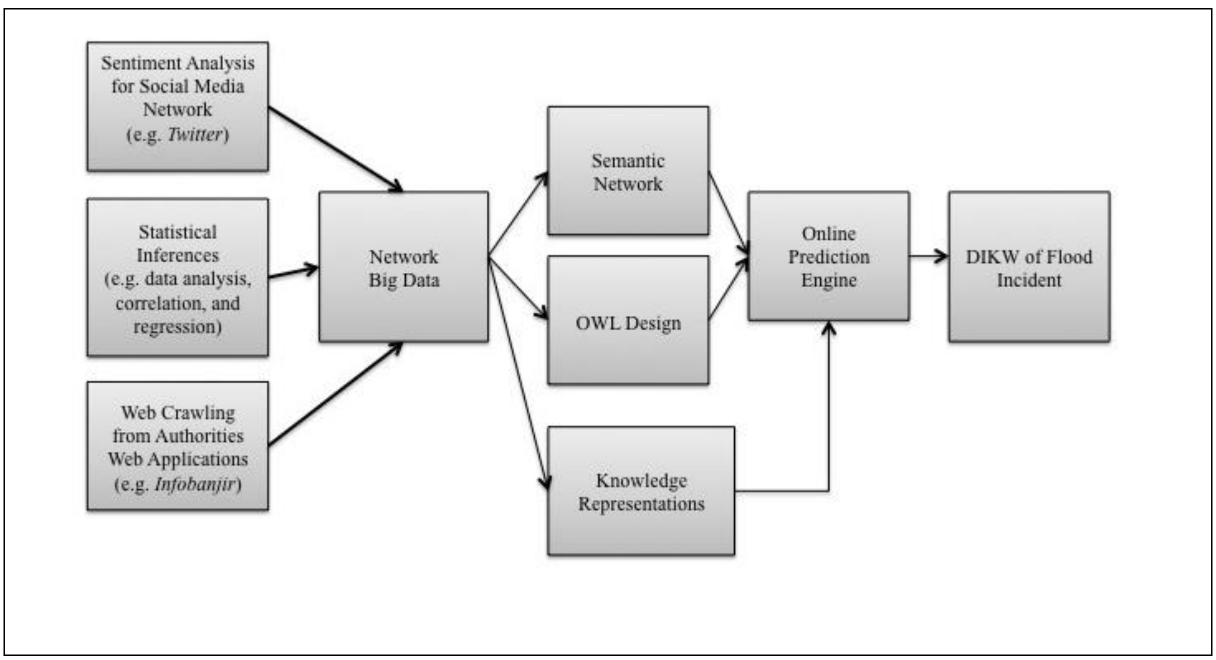

Fig. 6. The Network Big Data Prediction Engine for DIKW of Flood Incident

Simultaneously, the graphical representation on how the DIKW of flood incident can be implemented is as illustrated in Fig. 7. This illustration is called a decision tree and depicted in accordance to the DIKW hierarchical level. The online prediction for big data should be able to produce a smart retrieval in this form of representation. Data is at the basic level, the most essential build of a decision sciences engineering. However, data is smart when the analysis is able to interpret the information and knowledge from the studied materials. The data is at wisdom level when more than just knowledge representations can be abstracted.

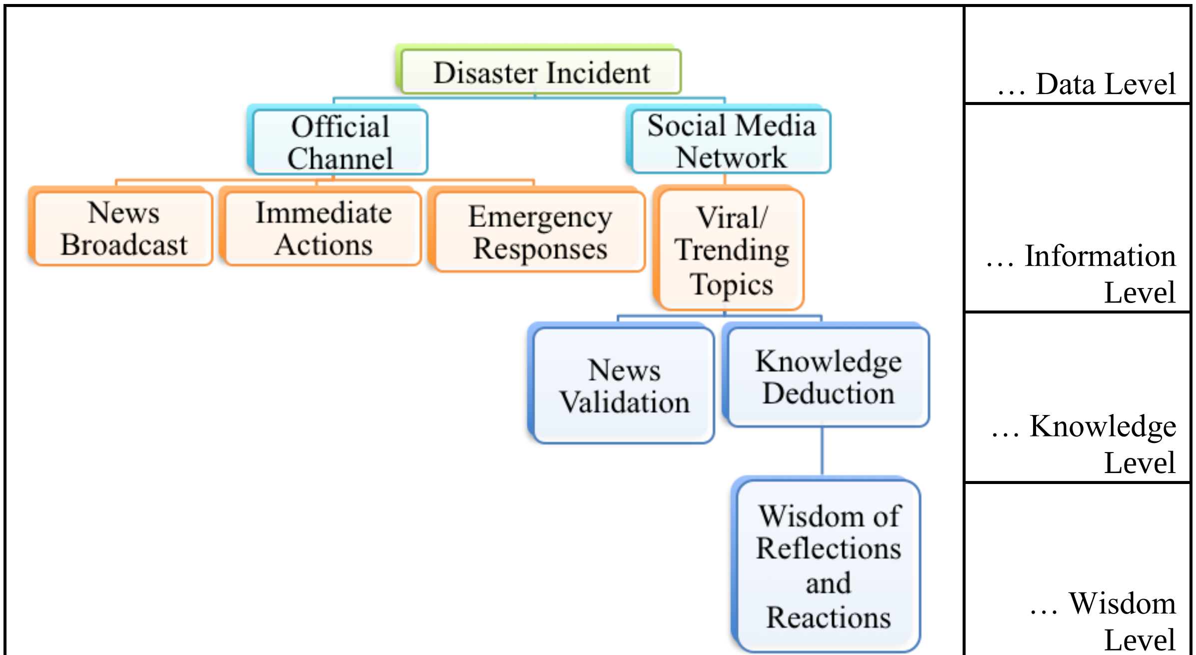

Fig. 7. The Decision Tree and DIKW Level of Abstraction for a Disaster Incident

13.6 Conclusion

The big data is always associated with its properties of 5Vs: volume, velocity, variety, veracity, and value. The challenge in this study was to utilize the network big data by performing a DIKW hierarchy to the product of its processes. As a comparison, the analytics approach was initially started with statistical testing method. From the statistical analysis, the study can define the hypotheses, the data to be manipulated, the operations such as correlation and regression and lastly the prediction by normal distribution forecasting method. However, this approach is further deliberated by defining the ontological relationships of the data and its propositional logic. The semantic network is defining the classes and subclasses of the research study. OWL design is chosen as the platform to design the parameters. The knowledge representations of the semantic web relationships are defined as the properties, sub-properties, and functional properties to the classes. At the end of the process, an online prediction is expected to run. This online prediction engine is meant to perform a DIKW hierarchical outputs for specific case study that is performed.

References:

[1] Aven, Terje. "A conceptual framework for linking risk and the elements of the data–information–knowledge–wisdom (DIKW) hierarchy." *Reliability Engineering & System Safety* 111 (2013): 30-36.

[2] Baskarada, Sasa, and Andy Koronios. "Data, information, knowledge, wisdom (DIKW): a semiotic theoretical and empirical exploration of the hierarchy and its quality

dimension." *Australasian Journal of Information Systems* 18, no. 1 (2013).

[3] Sharma, Nikhil. "The origin of DIKW Hierarchy." *Go. webassistant. com* 11 (2004).

[4] Jifa, Gu, and Zhang Lingling. "Data, DIKW, Big data and Data science." *Procedia Computer Science* 31 (2014): 814-821.

[5] Ackoff, Russell L. "From data to wisdom." *Journal of applied systems analysis* 16, no. 1 (1989): 3-9.

[6] Zeleny, Milan. "Management support systems: towards integrated knowledge management." *Human systems management* 7, no. 1 (1987): 59-70.

[7] Yusoff, Aziyati, Intan Shafinaz Mustafa, Salman Yussof, and Norashidah Md Din. "Green cloud platform for flood early detection warning system in smart city." In *Information Technology: Towards New Smart World (NSITNSW), 2015 5th National Symposium on*, pp. 1-6. IEEE, (2015).

[8] Wolfram, Stephen. *A new kind of science*. Vol. 5. Champaign: Wolfram media, 2002.

[9] Rowley, Jennifer E. "The wisdom hierarchy: representations of the DIKW hierarchy." *Journal of information science* (2007).

[10] Batra, Surinder. "Big Data Analytics and its Reflections on DIKW Hierarchy." *Review of Management* 4, no. 1/2 (2014): 5.

[11] Katal, Avita, Mohammad Wazid, and R. H. Goudar. "Big data: issues, challenges, tools and good practices." In *Contemporary Computing (IC3), 2013 Sixth International Conference on*, pp. 404-409. IEEE, (2013).

[12] Singh, Sachchidanand, and Nirmala Singh. "Big data analytics, 2012 Internati-onal Conference on Communication." *Information Computing Technology (ICCICT 2012)* 4 (2012).

[13] Sekaran, Uma, and Roger Bougie. *Research Methods for Business: A Skill-building Approach*. Chichester: Wiley, (2010).

[14] Yusoff, Aziyati, Norashidah Md Din, Salman Yussof, and Samee Ullah Khan. "Big data analytics for Flood Information Management in Kelantan, Malaysia." In *2015 IEEE Student Conference on Research and Development (SCOReD)*, pp. 311-316. IEEE, (2015).

[15] Asia-Pacific Economic Cooperation (APEC). *ANSSR: Enhancing the Quality and Relevance of Technical and Vocational Education and Training (TVET) for Current and Future Industry Needs*. APEC Publications, (2014).

[16] Navidi, William Cyrus. *Statistics for engineers and scientists*. Vol. 1. New York: McGraw-Hill, (2006).

[17] Shroff, Gautam. *The Intelligent Web: Search, smart algorithms, and big data*. OUP Oxford, (2013).

[18] Zakaria A. S., "*The Geomorphology of Kelantan Delta (Malaysia)"* (Catena 2) pp 337-349, (1975).

[19] L. Nielsen, "*Hadoop: The Engine That Drives Big Data*", New Street Communications LLC, New York, (2013).

[20] V. Mayer-Schönberger, and Kenneth Cukier, "*Big data: A revolution that will transform how we live, work, and think*," Houghton Mifflin Harcourt, (2013).

[21] Hurwitz, Judith, Marcia Kaufman, and Adrian Bowles. *Cognitive computing and big data analytics*. John Wiley & Sons, (2015).

[22] Ohlhorst, Frank J. *Big data analytics: turning big data into big money*. John Wiley & Sons, (2012).

[23] John F. Sowa, "Semantic Networks". In Stuart C Shapiro. *Encyclopedia of Artificial Intelligence*. Retrieved 2008-04-29. (1987)

[24] Guarino, Nicola, Daniel Oberle, and Steffen Staab. "What is an Ontology?." In *Handbook on ontologies*, pp. 1-17. Springer Berlin Heidelberg, (2009).

[25] Aronson, Jaye, T. Liang, and E. Turban. "Decision support systems and intelligent systems." *Yoyakarta: Andi* (2005): 24.

[26] McGuinness, Deborah L., and Frank Van Harmelen. "OWL web ontology language overview." *W3C recommendation* 10, no. 10 (2004): 2004.

[27] Yusoff, Aziyati, Norashidah Md Din, Salman Yussof, and Samee Ullah Khan. "The Semantic Network of Flood Hydrological Data for Kelantan, Malaysia." In *IOP Conference Series: Earth and Environmental Science*, vol. 32, no. 1, p. 012021. IOP Publishing, (2016).

[28] Thangaraj, M., and G. Sujatha. "An architectural design for effective information retrieval in semantic web." *Expert Systems with Applications* 41, no. 18 (2014): 8225-8233.