Specifying Architecture of Knowledge Graph with Data Graph, Information Graph, Knowledge Graph and Wisdom Graph

Yucong Duan
State Key Laboratory of Marine Resource Utilization in the South China Sea, College of Information and Technology, Hainan University
duanyucong@hotmail.com

Lixu Shao
State Key Laboratory of Marine Resource Utilization in the South China Sea, College of Information and Technology, Hainan University
751486692@qq.com

Gongzhu Hu
Department of Computer Science
Center Michigan University
Michigan, USA
hu1g@cmich.edu

Zhangbing Zhou
China University of Geoscience (Beijing), Beijing, China
zhangbing.zhou@gmail.com

Quan Zou
College of Computer Science
Tianjin University
zouquan@tju.edu.com

Zhaoxin Lin
School of business
Iowa State University, USA
zxlin@iastate.edu

***Abstract*—Knowledge graphs have been widely adopted, in large part owing to their schema-less nature. It enables knowledge graphs to grow seamlessly and allows for new relationships and entities as needed. Knowledge graph has become a powerful tool to represent knowledge in the form of a labelled directed graph and to give semantics to textual information. A knowledge graph is a graph constructed by representing each item, entity and user as nodes, and linking those nodes that interact with each other via edges. Knowledge graph has abundant natural semantics and can contain various and more complete information. Its expression mechanism is close to natural language. However, we still lack a unified definition and standard expression form of knowledge graph. We propose to clarify the expression of knowledge graph as a whole. We clarify the architecture of knowledge graph from data, information, knowledge, and wisdom aspects respectively. We also propose to specify knowledge graph in a progressive manner as four basic forms including data graph, information graph, knowledge graph and wisdom graph.**

Keywords—knowledge graph, data, information, knowledge, wisdom

I. Introduction

There are different kinds of discrete data in the real world we live in. The data cannot be used if they exist only in the discrete form. However, this is not worth worrying as we can simply make the data meaningful by giving a specific environment. At this point, the data can get real meaning in this specified environment we supplied. The data are processed to be useful and presented to us in the form of information, then we can get a lot of fragmented expressions. For example, "Li Ming's height is 177cm" is the information is capable of being used to answer questions about Li Ming's height. Obviously, getting information is just a start. With these fragmented expressions, that is, the conception "information" we mentioned above, we can combine multiple information to answer more complex questions about how to do it. By abstracting and converting the information and the data in a given context and the application of data and information [2], "Knowledge" shows up. At this point, by defining the attributes and classification summary, we can digging the information from the data, and we can summarize knowledge from the information. Furthermore, comprehensive knowledge of the same category can be use of making favorable judgments, precisely predicting, and smartly planning. Obviously, the utilization of vested knowledge is beyond its literal meaning of the category, which is what we say, "wisdom". In this paper, we propose to specify the architecture of knowledge graph in four aspects including data graph, information graph, knowledge graph and wisdom graph as well as the practical significance of the analysis of the four conceptions. Fig.1 shows the conversion from data to wisdom. Data existing as discrete elements have no semantics. Information is data after procession of conceptual mapping and relational connection. The user access to information after filtering the valuable information and internalized those information into knowledge. When information is adequately assimilated, it produces knowledge, modifies the individual's mental store of information and benefits his development and that of the society in which he lives.

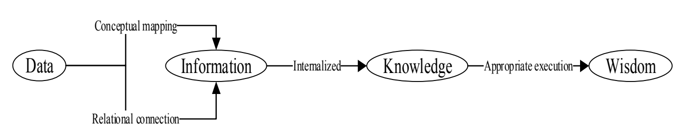

Fig. 1. Conversion process from data to wisdom

In the rest of this paper, we firstly elaborate representations of data graph, information graph, knowledge graph and wisdom graph in Section 2, 3, 4 and 5respectively. Then we describe the progressive relationship among data, information, knowledge and wisdom in Section 6. The related works are elaborated in Section 7. And we conclude our work in Section 8.

II. Representation of Data Graph

Data are the symbolic representation of observable properties of the world. Data are obtained by observing the basic individual items of numbers or other information, but on their own, without context, they have no meaning. Storing of data does not change the data itself, but it has many expression forms. As Fig.2 shows, data can be organized in many different types of data structures, including arrays, stacks, list links and so on. The data can be structured, semi-structured and unstructured, relational or non-relational. Data structures can store a great deal of data in many different types, including numbers, strings and even other data structures. Generally, data can be represented as many discrete elements originally [8]. Fig. 3 shows a series of original discrete data points and that we can use array, linked list, tree, graph as well as the combination of these four structures to represent data respectively. Original discrete data points have no meaning without context. For example, the value 120 can be clinical measurement such as heart rate and it can also indicates the telephone number of the emergency center. Cutting them from the specific context of situations, we cannot determine what it means for sure. We use a collection D {d1, d2, …, di, …, dn} to represent the data sets where di indicates a discrete element. For example, if we input a collection of a series of discrete elements describing risk assessment of software engineering it can be denoted as D (risk). We cannot understand the specific meaning of each element without context and the internal relationship of these elements.

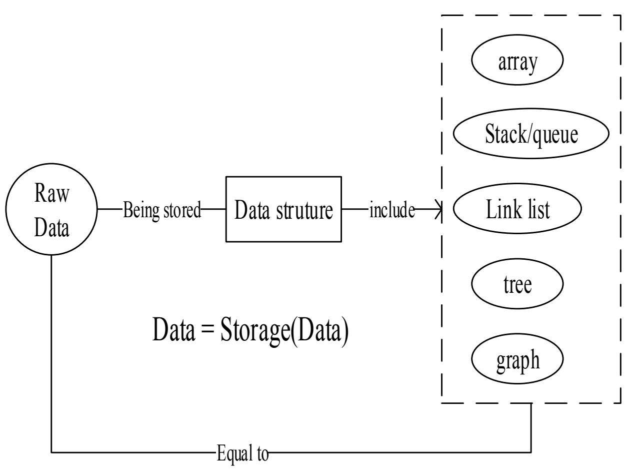

Fig. 2. Storage of data does not change data itself

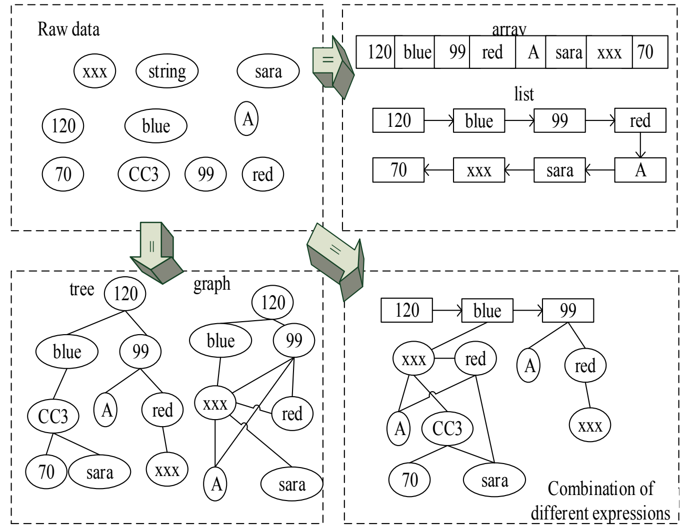

Fig. 3. Illustration of data and data expression

III. Representation of Information Graph

Information is data that has been given meaning by way of relational connection. This "meaning" can be useful, but does not have to be. Items of information include elements of information and relations between the elements of information. The elements of information are displayed as nodes and relations are displayed as lines on the information graph. Information embodies the understanding of a relationship of some sort and the essence of information phenomenon has been characterized as the occurrence of a communication process that takes place between the sender and the recipient of the message. The conceptual mapping of different concepts and relationships is called concept mapping [3]. Giving two concepts O1, O2 and their associated sets C1 & R1 and C2 & R2, conceptual mapping is to identify potential pairs: (c1, c2) or (r1, r2), where c1 ∈ C1, c2 ∈ C2 , r1∈R1 and r2∈R2. In this way, concept c1 and relationship r1 can each be translated into instance c2 and instance r2 while preserving their original meaning. We define a conceptual mapping function F acting on two concepts C1 and C2. The similarity function is defined in two concepts C1 and C2, and a value between 0 and 1 is calculated to indicate the similarity between C1 and C2. The logical representation of the concept Ci is L (Ci) where the function F is evaluated as a similarity representation of the logical representation between C1 and C2. The function S is used to define the logical similarity evaluation on L (C1) and L (C2).

$$F(C1, C2) = S\big(L(C1), L(C2)\big) \qquad (1)$$

There is a need to apply the conceptual semantics of professionals from different aspects. Correspondingly, different semantics can be represented by independent logical statements, and function F can be defined exactly as follows:

$$F(C1, C2) = \sum_{i} w(i) * S\big(L_i(C1), L_i(C2)\big) \qquad (2)$$

where i indicates a kind of features and its logical expression is defined as L_i. w(i) is an application collaborative function that is applied to i (determined by the application professional) to measure the importance of each i in evaluating the similarity of C1 and C2. Fig. 4 illustrates that a series of raw data points can be converted to information through conceptual mapping. Relationships between the information obtained through conceptual mapping are consistent with relationships between the original concepts. On the information graph, there is a simple combination relationship between data points. The contextual relevance of the information is limited, and in different contexts we can establish different classification and combination rules. In Fig. 5 we can recognize that there are three kinds of data including risk factors, combination type of these factors and the corresponding probability of each combination type through conceptual mapping. Then we can have a more complete description of risks that can denoted as D (A, C, P) and store the description in relational database. A indicates factors that may lead to risk. C indicates the combination type of these factors and P indicates the corresponding probability of each combination type. A= {system type, developing experience, software and hardware equipment}. C includes eight combination types of the three factors and P includes the corresponding eight probabilities.

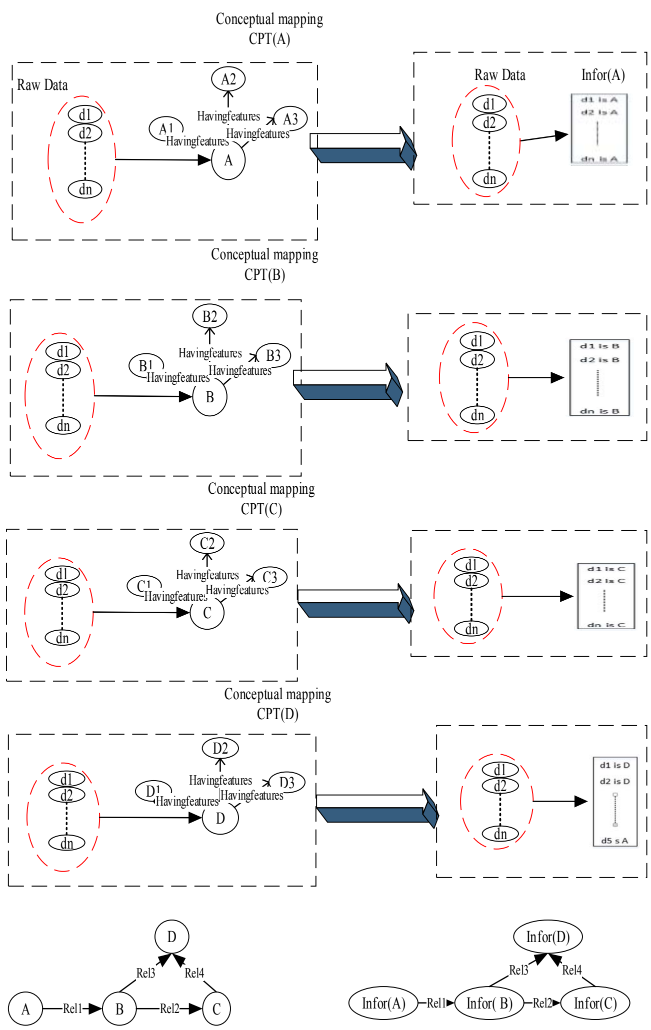

Fig. 4. Conceptual mapping from data to information

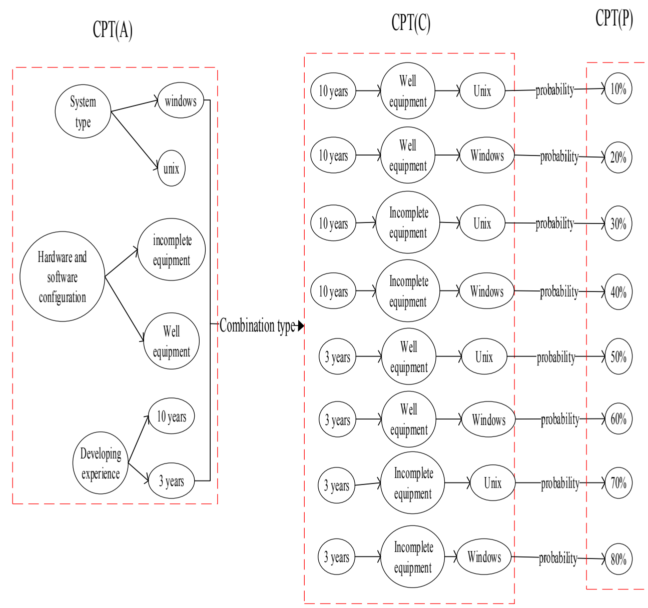

Fig. 5. Information graph after relational connection

IV. Representation of Knowledge Grpah

Knowledge is information that is structured and organized as a result of cognitive processing and validation. Information is a necessary medium or material for eliciting and constructing knowledge. Knowledge is capable of being gained by learning. For knowledge to be passed on entails encoding knowledge into information and decoding again into knowledge. Knowledge and information are not the same, but have as symbiotic relationship [9]. Knowledge may be explicit for instance written guidelines and implicit such as people's experience and intuition. The purpose of knowledge is to better our lives. In the context of risk assessment, the purpose of knowledge is to reduce the risk rate or to avoid risks as much as possible for the enterprise and all its stakeholders. Knowledge representation and reasoning formalism can also expression problems to be solved concerning facts and general knowledge represented [11]. For instance, one may ask with what kind of languages does Mike speaking. Answering such questions requires descriptive knowledge but also reasoning capabilities.

A. Abstraction on Knowledge Graph

Data and information are complex, from which we extract valuable as a knowledge. Thereby we are capable of reducing the available information capacity. When stakeholders obtain the description of risk they are able to screen out the valuable information and preserve those information as knowledge. As for the example shown in Fig.5, decision maker will choose a program with a lower rate of risk as Fig.7 shows. Page rank algorithm works by counting the number and quality of links to a page to determine a rough estimate of how important the website is. The underlying assumption is that more important websites are likely to receive more links from other websites [15]. We adopt the idea of Page rank algorithm to filter useless information and retain valuable information. Relevance is measured as the probability that retrieved resource actually contains those relations whose existence was assumed by the user at the time of query definition. As Fig.7 shows, we give each raw data contained in information a certain initial weight. Ranks of information, denoted as $R_{(infor)}$, can be calculated according to (3):

$$R_{(Infor)} = w_{A_i} * w_{B_j} * w_{C_k} * \dots * w_{N_n}, \quad i,j,k,n \subset \{1,2 \dots n\} \tag{3}$$

where A indicates concepts and Ai indicates the raw data elements of concept A. W_{Ai} represents the weight of element Ai. After calculating ranks of information, we can filter out information that does not meet users' query definition.

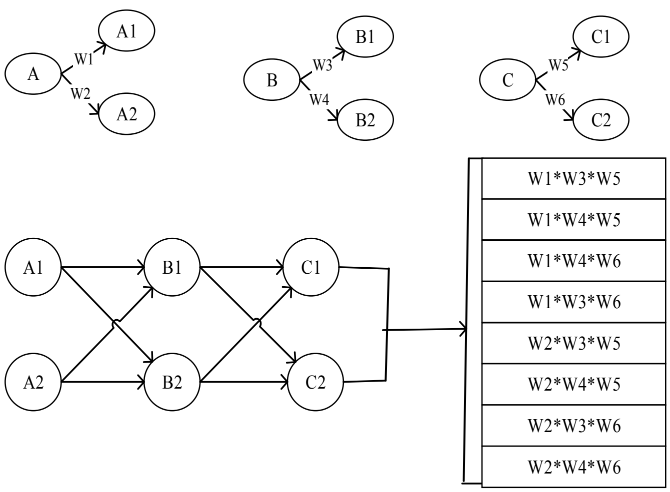

Fig. 6. Calculating ranks of information

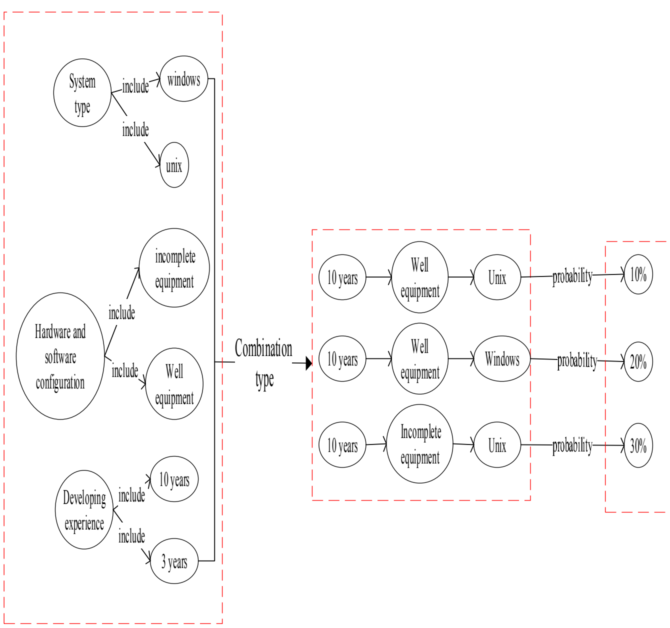

Fig. 7. Abstraced knowledge graph after filtering useless information

B. *Transformation on Knowledge Graph*

With knowledge stakeholders can make more correct decision. The context of knowledge graph can be created. Knowledge graph can provide an open knowledge access interface and to a certain extent it reflects the real world of inter-entity relations. The graph structure of knowledge graph in Fig. 8 is not restricted by form. Knowledge graph can express abundant natural semantics and can supplement related information among terms. The graph-based nature of knowledge graph makes possible a linkage to other graphs thus resulting in an easy integrating of multiple kinds of information and an enhancement in integrity of information. By exploring the graph, new connections and commonalities between items and users can be discovered and exploited.

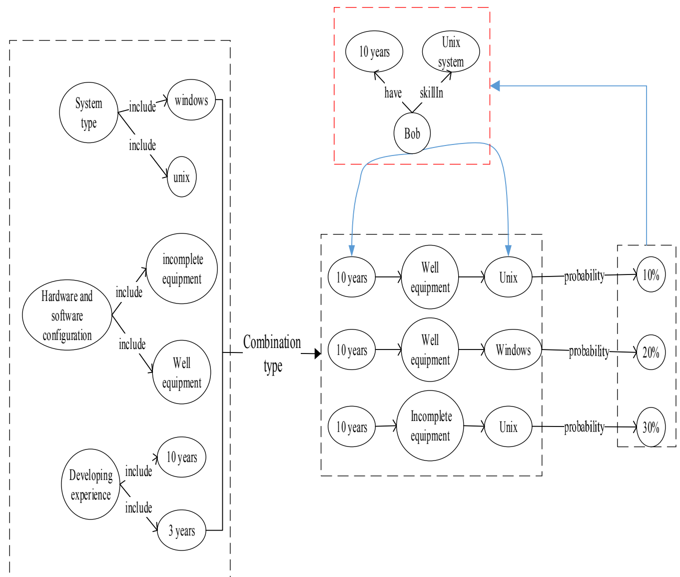

Fig. 8. Supplementing semantic terms on knowledge graph

V. Representation of Wisdom Graph

Wisdom is an extrapolative process which includes knowledge in an ethical and moral framework. Wisdom is the process by which we can discern right from wrong and good from bad. With wisdom we can judge from limited to infinity, from known to unknown. Wisdom is the capacity to put into action the most appropriate behavior, taking into account what is known (knowledge) and what does the most good (ethical and social considerations). Many informed people know what to do, quite a few knowledgeable experts know how to do it, but only a few wise persons know and can fully explicate why it should be done. In line with these ideas the following metaphor applies in Fig. 9, data: "know-who/when/where", information: "know-what", knowledge: "know-how" and wisdom: "know-why". Wisdom as the ultimate unit of cognition is the result of hierarchical processing of data, information, concept, and knowledge. Knowledge is "knowing how" to do something, wisdom is "knowing why, what and how" to do something. Wisdom also extends to the application of knowledge in action. A simplistic representation of the relationship between wisdom and knowledge is captured in the following expression: Wisdom = Knowledge + Ethics +

Action. In Table 1 we take the whole process of risk analysis of software engineering as an example, and list the relations of data, information, knowledge and wisdom in the process of risk analysis. At the initial stage of the risk assessment, we collected some data about risk assessment, which can be stored in arrays, stacks, lists, and so on. Based on the data we collect, we can get descriptive information about risk through conceptual mapping and relational connection. And then according to the concept of classification we store information in the relational database. Stakeholders can make a favorable decision after gaining this risk description, and ultimately wisdom can help stakeholders make future planning and forecasting.

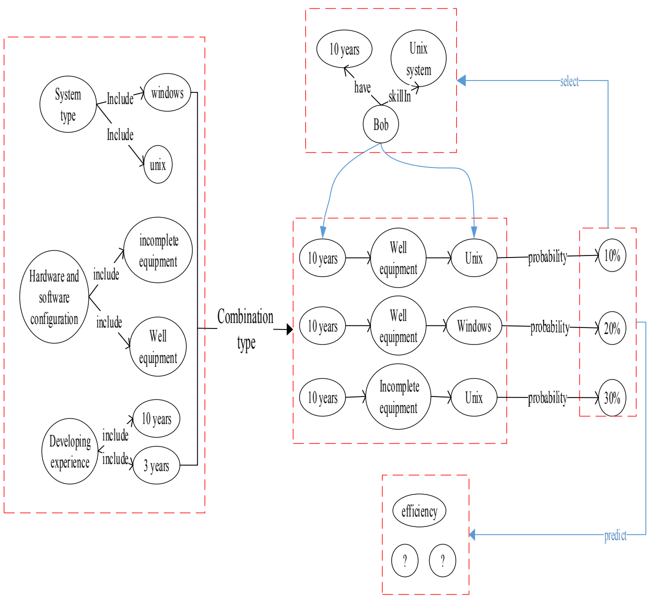

Fig. 9. Using wisdom graph to predict unknown elements

TABLE I. EXAMPLE EXPLANATION IN DATA, INFORMATION, KNOWLEDGE AND WISDOM ASPECTS

Aspects	Semantic load	Expression
Data	Input of risk assessment	Array, list link, stack, queue, tree, graph
Information	Risk description	Relational database
Knowledge	Understanding the description and make decision	Nodes and edges with semantic relations
Wisdom	(understand why) the ability to use results of the analysis in the right way	(frame or stylish expression)

VI. PROGRESSIVE RELATIONSHIP OF DATA, INFORMATION, KNOWLEDGE AND WISDOM

In Fig. 10, we show the progressive relationship between data, information, knowledge, and wisdom through a pyramid form. Data is one of the primary forms of information. It basically includes recordings of transactions or events that will be used for exchange between human or even with the machine. Thus, unless the user understands the context in which the data is collected, the data does not make sense. A word, a number or a symbol can be used to describe a business result. It is the context that gives data meaning, and this meaning makes data informative. Information extends the concept of data in a broader context. Therefore, information includes data but it also includes all the information that a person associates as a member of a social organization in a given physical environment. Information like data is passed by symbol. These symbols have complex structures and rules. Information has various forms, such as writings, statements, statistics, charts or diagrams. When an individual accepts and retains information as a true understanding of reality and an effective explanation of reality, the information becomes personal knowledge. On the contrary, the organization or social knowledge exists when it is accepted by the consensus of a group of people. Common knowledge does not need to be shared by all members, and in fact it is accepted by a group of insiders that can be considered a sufficient condition. This is also the "public domain" knowledge of the real [10]. Knowledge is a step further on the scale. It involves understanding and ability to make use of the data and information to answer questions, to solve problems, to make decisions and so on. As the human mind uses this knowledge to choose between alternatives, behavior becomes wise. Finally, when values and commitment guide intelligent behavior, behavior may be said to be based on wisdom. The level of wisdom includes all the required components such as data, information, and knowledge to make wise decisions.

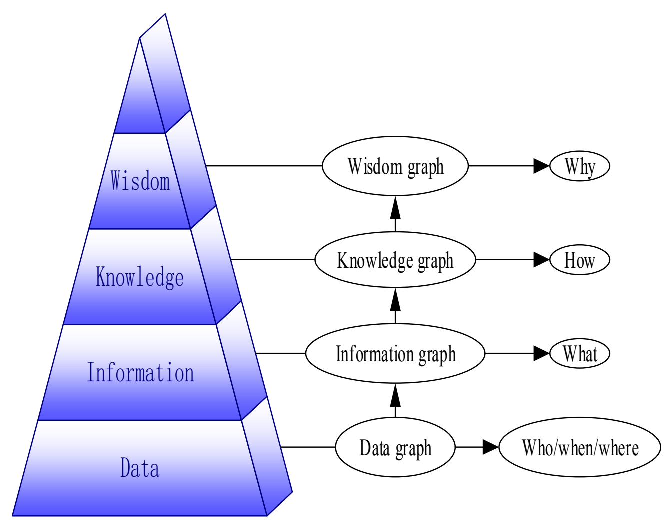

Fig. 10. Relationship among data, information, knowledge and wisdom

VII. RELATED WORKS

Knowledge representation is a critical topic in AI, and currently embedding as a key branch of knowledge representation takes the numerical form of entities and relations to combine the statistical models. However, most embedding methods merely concentrate on the triple fitting and ignore the explicit semantic expression, leading to an uninterpretable representation form [12, 13]. Traditional embedding methods do not only degrade the performance, but also restrict many potential applications. In [7] the authors proposed a semantic representation method for knowledge graph, which imposes a two-level hierarchical generative process that globally extracts many aspects and then locally assigns a specific category in each aspect for every triple. Because both the aspects and categories are semantics-relevant, the collection of categories in each aspect is treated as the semantic representation of this triple. In [14] the authors proposed to represent knowledge in logical, philosophical, and computational foundations. As an instance of guidance in the practice of Value Driven Design, in [16] the authors proposed a systemic formalization from the value calculation to the design quality measurement which binds the modification and change on the design artifacts with the business value strategy through a framework of managed quality properties in a service design process.

VIII. CONCLUSION

Knowledge graph has been widely adopted these years, but the expression of knowledge graph is usually limited to the form of triples. We are increasingly aware of the semantic functions of knowledge graph. In this paper, we elaborate the relationships among data, information, knowledge, and wisdom, with the aim of clarifying the expression of knowledge graph from the four levels data, information, knowledge, and wisdom. And we propose to specify the architecture of knowledge graph in the four aspects of "data graph, information graph, knowledge graph and wisdom graph". For users with complex information needs, they can be allowed to express their needs by proposing natural language questions. We can use the data graph to answer questions asked by "Who / When / Where". Data are meaningless in the absence of a given context. Information is a combination of discrete data that gives an answer to the question directed by "what". Knowledge is an effective combination of abstracted and transformational information and capable to answer questions guided by “what”. Wisdom is the ability to criticize or act in a given situation. Wisdom can provide the answer to "why" questions. Our work lays the foundation for a survey from data to wisdom. In the next stage, we will deal with data, information, knowledge and wisdom on the same background and different backgrounds of the 5Ws problem and explore more accurate expression of knowledge graph.

ACKNOWLEDGMENT

This paper is supported by National Natural Science Foundation of China under Grant (No.61363007, No.61662021 and No. 61662019) and Natural Science Foundation of Hainan under Grand No. ZDYF2017128.*refers to corresponding author.

REFERENCES

[1] Chaim Zins. Conceptual approaches for defining data, information, and knowledge [J]. Journal of the American society for information science and technology, 2007, 58(4):479-493.

[2] Bellinger G, Castro D, Mills A. Data, information, knowledge, and wisdom [J]. 2004.

[3] Nadarajan G, Chenburger Y H. An Ontology-Based Conceptual Mapping Framework for Translating FBPML to the Web Services Ontology[C]// IEEE International Symposium on CLUSTER Computing and the Grid. IEEE Xplore, 2006:8 pp.

[4] Nadarajan G, Chenburger Y H. An Ontology-Based Conceptual Mapping Framework for Translating FBPML to the Web Services Ontology[C]// IEEE International Symposium on CLUSTER Computing and the Grid. IEEE Xplore, 2006:8 pp.

[5] Lamberti F, Sanna A, Demartini C. A Relation-Based Page Rank Algorithm for Semantic Web Search Engines[J]. IEEE Transactions on Knowledge & Data Engineering, 2008, 21(1):123-136.

[6] Norta A, Ma L, Duan Y, et al. eContractual choreography-language properties towards cross-organizational business collaboration [J]. Journal of Internet Services and Applications, 2015, 6(1): 8.

[7] Xiao H, Huang M, Zhu X. Knowledge Semantic Representation: A Generative Model for Interpretable Knowledge Graph Embedding[J]. 2016.

[8] Johannessen T V, Fuglseth A M. The effectiveness of data presentation formats: An exploratory study[C]// Nokobit 2014, norsk konferanse for organisasjoners bruk av IT. 2014.

[9] Lamberti D M, Wallace W A. Intelligent Interface Design: An Empirical Assessment of Knowledge Presentation in Expert Systems[J]. Mis Quarterly, 1990, 14(3):279-311.

[10] Syed A, Shah A. Data, Information, Knowledge, Wisdom: A Doubly Linked Chain?[C]// International Conference on Information & Knowledge Engineering, Ike 2006, Las Vegas, Nevada, Usa, June. DBLP, 2006:270-278.

[11] Chein M, Mugnier M L. Graph-based knowledge representation: computational foundations of conceptual graphs[M]. Springer Science & Business Media, 2008.

[12] Guu K, Miller J, Liang P. Traversing Knowledge Graphs in Vector Space [J]. Computer Science, 2015.

[13] Minervini P, Fanizzi N, D'Amato C, et al. Scalable Learning of Entity and Predicate Embeddings for Knowledge Graph Completion[C]// IEEE, International Conference on Machine Learning and Applications. IEEE, 2015:162-167.

[14] Sowa J F. Knowledge representation: logical, philosophical, and computational foundations[M]. Pacific Grove: Brooks/Cole, 2000.

[15] Lamberti F, Sanna A, Demartini C. A Relation-Based Page Rank Algorithm for Semantic Web Search Engines[J]. IEEE Transactions on Knowledge & Data Engineering, 2008, 21(1):123-136.

[16] Duan Y, Duan Q, Hu G and Sun X. Refinement from service economics planning to ubiquitous services implementation[C]//Computer and Information Science (ICIS), 2016 IEEE/ACIS 15th International Conference on. IEEE, 2016: 1-6.