IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 16, 2020, accepted February 26, 2020, date of publication February 28, 2020, date of current version March 11, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2977136

KnowIME: A System to Construct a Knowledge Graph for Intelligent Manufacturing Equipment

HEHUA YAN[1], JUN YANG[2], AND JIAFU WAN[2]

[1]School of Electrical Technology, Guangdong Mechanical and Electrical Polytechnic, Guangzhou 510515, China
[2]School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou 510641, China

Corresponding author: Jiafu Wan (jiafuwan_76@163.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFE0101000, in part by the Key Program of National Natural Science Foundation of China under Grant U1801264, in part by the Key Areas Research and Development Program of Guangdong Province, China, under Grant 2019B010150002 and Grant 2019B090919002, and in part by the Key Program of Natural Science Foundation of Guangdong Province, China, under Grant 2017B030311008.

ABSTRACT With the development of a new generation of information technology, such as big data and cognitive intelligence, we are in the postmodern era of artificial intelligence. Currently, the manufacturing industry is in the critical period of transitioning to smart manufacturing, but the cognitive capabilities of devices in smart factories are still scarce. Knowledge Graph (KG) is one of the key technologies of cognitive intelligence, which opens a new path for the horizontal integration of intelligent manufacturing. Therefore, this paper proposes and builds a manufacturing equipment information query system based on KG. Firstly, a large amount of heterogeneous data that contains vast devices information is obtained from the network. Secondly, the conditional random fields (CRF) algorithm is used to extract the entity name, product place, and company name of the device, and then the relationship between the device entities is identified by calculating the similarity and Chinese syntax analysis. In the validation section, we use to the map of Neo4j graph database, when we input a name of a device in the search box, the system can return a relational graph node. In addition, the shortest path optimization algorithm is used to calculate the similarity between nodes in the search process to achieve the recommendation of similar node information.

INDEX TERMS Intelligent manufacturing equipment, knowledge graph, Neo4j, CRF, syntactic analysis.

I. INTRODUCTION

The most fundamental purpose of Industry 4.0 is to transform traditional industrial manufacturing with advanced information technology [1], [2]. The core issue involved is information integration, which consists primarily of enterprise data value chains in the vertical and horizontal dimensions. Horizontal integration focuses on the information integration of the industrial chain to optimize the processes of R & D, procurement, manufacturing, and services among enterprises [3]. Meanwhile, vertical integration focuses on the integration from device to cloud within the enterprise, which significantly improves manufacturing data utilization and product on-site delivery efficiency. However, from the perspective of companies with low-level manufacturing equipment and management systems, they urgently need more information about products and equipment in horizontal integration, such as equipment categories, prices and suppliers [4]–[6].

With the development and application of new technologies such as the Internet of Things (IoT) [7]–[9], cloud computing [10], big data, artificial intelligence [11], [12] and mobile communications [13], the horizontal integration of manufacturing information resources of smart factories has been greatly developed. Qi *et al.* [14] proposed a supportive design and tools for scalable, modular integrated manufacturing integration and standardization. Their method effectively implemented enterprise application integration based on extensible language and Web service technology. Qi *et al.* [15] reviewed the application of big data and digital twin in manufacturing, including their concepts and applications in product design, production planning, manufacturing and predictive maintenance. With the development of intelligent manufacturing, a large amount of data is generated every day, and the semantics of data plays an important role in the extraction and application of manufacturing information.

The associate editor coordinating the review of this manuscript and approving it for publication was Zheng Liu.

The importance of data for intelligent manufacturing is self-evident, and in the construction of smart factories, there is a problem of how to effectively integrate the information of products and equipment. Knowledge map (KG) is a structured semantic knowledge base that describes the concepts and relationships in the physical world in symbolic form [16]. With the help of KG, you can integrate ''knowledge'' with different equipment information, and obtain the required information from the manufacturing equipment and its attributes. In this way, not only can all device information pass through the entire life cycle of the device product, but this comprehensive information can be applied into improving the replacement of devices in the manufacturing process and achieving rapid iteration of the system.

Actually, KG integrates heterogeneous data (including structured, semi-structured and unstructured data) from different sources (network, manufacturer, etc.) to form a knowledge base of graph, where nodes represent entities and edges represent the relationship [17]. As a prominent case in the industry, Google KG uses graph data structures to show the entities and relationships of the physical world [18]. And a knowledge mapping system for the field of education was proposed to explore the relationship behind educational entities in [19]. To realize the information reconfigurability of intelligent manufacturing equipment, Wan *et al.* [20] proposed a method of intelligent manufacturing resource reconstruction based on a knowledge base. Based on the above research, this paper uses KG to construct the architecture of manufacturing equipment information knowledge system, which mainly includes data acquisition, data processing and knowledge mapping.

The highlights of the paper are as follows:

- Based on multi-source heterogeneous data, KG are used to build information integration systems for manufacturing equipment such as lathes, conveyors and robots.
- The intelligent manufacturing equipment entity attribute relationship has a certain particularity compared with the general KG. The CRF algorithm is used to complete the extraction task of the device name entity, and the unsupervised syntax analysis method is used to complete the device relationship identification.
- Save the processed data to the database and use Neo4j to build the KnowIME (KG's Intelligent Manufacturing Equipment) information system.

The rest of this article is organized as follows. Section II introduces related works. The system architecture is proposed in section III. Section IV gives the algorithm and manual operation of the extraction and annotation of device entities and the relationships. In Section V, the query and recommendation of the KnowIME information system is constructed. Section VI discusses the limitations and future work of this paper. And finally, Section VII concludes the paper.

II. RELATED WORKS

There are some generalized KGs, such as Freebase, Reverb, and Microsoft's Probase. KG is divided into general-purpose KGs and dedicated KGs by application domain. The existing general KGs show their advantages in supporting many applications, which usually include semantic search (for example, Google's knowledge and IBM's Watson) [21]. By constructing a product KG, A new way for interpreting product features and functions is provided, and consumers can learn more about the details of product production [22]. Kim [23] proposed a knowledge representation framework based on semantic hypergraphs to support knowledge sharing in product development. The KG construction process generally includes the following steps: entity extraction, unit relationship extraction, and structured display [24]. Z. Liu *et al.* [25] presented the Entity-Duet Neural Ranking Model, which uses to the two components are learned end-to-end, making the Entity-Duet Neural Ranking Model (EDRM) a natural combination of entity-oriented search and neural information retrieval. The above methods or models significantly improve the application and generalization of KG on product information, but few studies focus on manufacturing field.

Entity extraction, one of the most critical steps in building a KG, is mainly used to extract concepts from structured and unstructured data. With the development of artificial intelligence, entity extraction methods based on deep neural networks have become more and more popular. Han *et al*. [26] compared the various cyclic units of the cyclic neural network and proposed that advanced gated repeat units be applied to sequence labeling. Chung *et al*. [27] proposed a two-way long-term and short-term memory network for named entity recognition. These advanced methods have improved the effect of entity extraction to a certain extent, but most of them require high-performance servers as the basis, and in a smart factory, it is difficult to meet the demand. Therefore, in this paper, the system mainly uses the CRF algorithm for entity extraction [28].

Another important aspect of building KG is the relationship extraction. After identifying the entity, it is necessary to determine whether any two entities in the same sentence constitute a previously defined or undefined semantic relationship. The relationship represents the semantic relationship between data, and the entities that established the connection will have practical significances. Huang *et al*. [29] have conducted in-depth research on remote monitoring methods, which can be combined with models to extract the relationships between entities. Quirk and Poon [30] used a piecewise convolutional neural network to train models for extracting entity relationships. Zeng [31] proposed a relationship extraction model based on sentence hierarchy. The completion and refinement techniques of a knowledge map can also be used to identify or predict undiscovered and lost relationships. These articles are very instructive for semantic relationship extraction, and they all have sufficient corpus for model training. However, due to the dispersion of manufacturing equipment information, there is no applicable corpus, so collecting sufficient equipment data is a prerequisite for relationship identification and entity extraction.

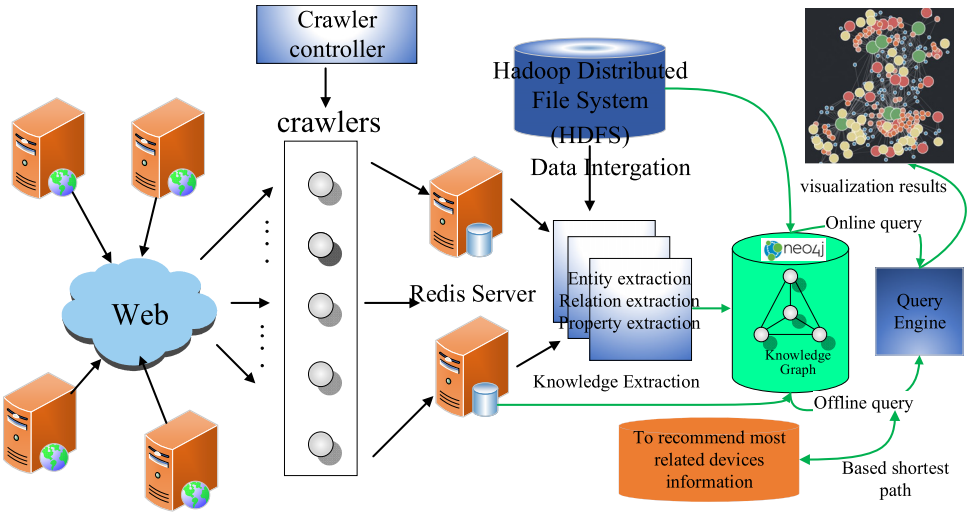


FIGURE 1. The knowIME information system architecture.

III. SYSTEM ARCHITECTURE

An architectural diagram of building a KnowIME information system is shown in Fig. 1. First, the unstructured data (e.g. text, images) and structured data (e.g. numerical data) of the intelligent manufacturing equipment are obtained from the Internet, Baidu Encyclopedia, and related intelligent manufacturing websites. The data is then saved to the Redis database along with the relational data. Second, knowledge entities and relationships from data came from multiple sources such as databases and Hadoop File System are extracted [32]. Then, the extracted entities and relationships are saved to specially formatted files and imported into the Neo4j non-relational database by means of the APOC tool and the load csv file. Among emerging technologies, NoSQL databases have the advantages of flexibility, scalability, efficiency, and the ability to handle large amounts of unstructured, semi-structured, and structured data. Finally, to improve the efficiency of users' query knowledge, we primarily complete the store of knowledge between the entities and relationships in the graph database and optimize the graph structure as much as possible.

In the system architecture, the focus is mainly on four parts: 1) Collect a large amount of relevant information about the intelligent manufacturing equipment. 2) Extract the relationship between device entities from structured and unstructured data, mainly using the probability map model of natural language processing (NLP). 3) Load the previously extracted entities and their connections into the Neo4j graph database to build the intelligent manufacturing device KG system. At the same time, the correlation graph algorithm such as breadth-first search (BFS) and depth-first search (DFS) is used to calculate the similarity of graph nodes, and Spark's GraphX is used to improve the efficiency of graph data information retrieval. And 4) Visualize the query information through the Echarts result graph, and the user can assist the production decision by querying the most relevant device information recommended [33].

IV. THE TECHNOLOGY OF BUILDING KG

As shown in Fig. 2, the proposed intelligent manufacturing equipment information system mainly includes two aspects, namely, entity extraction in the field of manufacturing equipment, and relationship extraction between entities in the equipment domain. Specifically, the data preparation phase is about obtaining data and cleaning data. Then, the operation of constructing the knowledge unit mainly includes the named entity information in the text and the relationship extraction between the unit entities. Structured display is the visualization process between the extracted entities and relationships using data visualization technology. And in the end, the shortest path algorithm is used to calculate the closest distance of the graph node to recommend relevant device information and provide the search service for users.

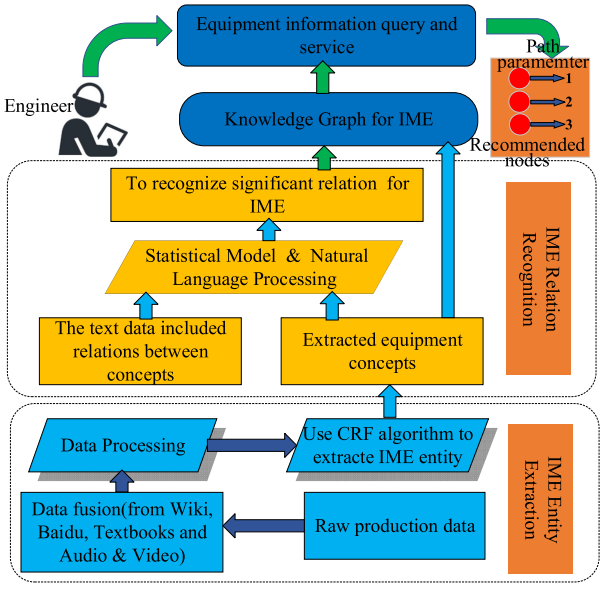


FIGURE 2. The process of built knowledge graph.

It is necessary to query related device information when a company or an individual is doing the underlying device integration. For example, to build an automated robot production line, engineers need a lot of different automation equipment, such as robots, machine tools, motors, and guide rails. In the past, the information they needed to query had to collect all the information about the device through the device's agent, which is one of the reasons why there are so many device agents. In contrast, our proposed system will help them save a lot of time.

A. DATA SOURCE AND PREPROCESSING

As mentioned earlier, in the KnowIME information system, the hope is that these nodes can cover as much information as possible. In this way, it is convenient for enterprises and engineers who do equipment integration to do retrieval. Therefore, the input data must be equipment-related data in the field of intelligent manufacturing, such as the manufacturer's product information data, Baidu Encyclopedia, and teaching laboratory equipment record data. This input data can contain different types, such as text, audio, and video, which must be converted into a data type that the computer can handle. For example, optical character recognition (OCR) technology can be used to process handwritten recorded data or printed text [34], and audio data processing involves NLP related techniques [35]. After the pre-processing steps of converting device-related data into a computer-readable format, our proposed system can extract the entities that make the device. This process is data fusion, which includes structured, semi-structured, and unstructured data from different sources. In this paper, based on the distributed storage of massive data by Hadoop, the memory-based Spark technology is used for data calculation and iteration to extract entities and update relationships for KG. And regarding the storage and query of data, the system uses NoSQL for unstructured and semi-structured data storage, using SQL for structured data storage.

The pre-processing of text data mainly uses the common methods of NLP, including the extraction of text and part-of-speech tagging. For the text data crawled from the web page, the sentence set required for the relationship extraction is obtained through the processing of word segmentation, removal of stop words, and part-of-speech tagging [36]. Concretely, firstly, web crawler technology is used to extract the text data in the web page and convert it into the required text data for storage. Secondly, the Jieba word segmentation tool is applied to complete word segmentation. Then, the stop word dictionary in Chinese is used to remove unnecessary words in the web page text. And finally, the required sentence is selected according to the token of the part of speech.

B. ENTITY EXTRACTION

Named entity recognition (NER) is one field of research in intelligent manufacturing. The early NER was achieved by manually compiling relevant rules, including the famous Porteus system developed by Grishman [37] and the FACILE system developed by Black [38]. Based on the above systems, through analyzing the morphological characteristics, context information, grammatical components, and wording rules of the named entity, the defined different types of named entities can be correctly identified.

The rule-based approach has great limitations in terms of generalization ability. As research has deepened, the NER method based on a statistical model and machine learning has become the mainstream method [39]. Some typical models are: Hidden Markov Model (HMM) [40], Support Vector Machine (SVM) [41], Maximum Entropy (ME) [42], and CRF [43]. Based on the above research of Chinese NER, this paper focuses on the problem of naming recognition of intelligent manufacturing equipment in the Chinese corpus and selects a machine learning method based on CRF model for the NER of intelligent manufacturing equipment.

Definition 1: Let $G = (V, E)$ be an undirected graph, V be a set of nodes, and E be a set of undirected edges. $Y = \{Y\_v \mid v \in V\}$, that is, each node in V corresponds to a random variable Y_v, which ranges from a possible set of tokens $\{y\}$. If the observation sequence X is used, then each random variable Y_v satisfies the following Markov characteristics:

$$p\left(Y_v|X, Y_\omega, \omega \neq v\right) = p\left(Y_v|X, Y_\omega, \omega \sim v\right) \tag{1}$$

where p represents the state transition probability, and $\omega \sim v$ denotes the adjacent points on the graph G. Then, (X, Y) is expressed as a conditional random field.

As shown in Fig. 3, the structure of the graph G may be arbitrary as long as certain conditional independence is described above the marked sequence. By modeling the sequence, a simple, ordinary chain structure diagram can be formed, and the nodes correspond to the elements in the mark sequence.

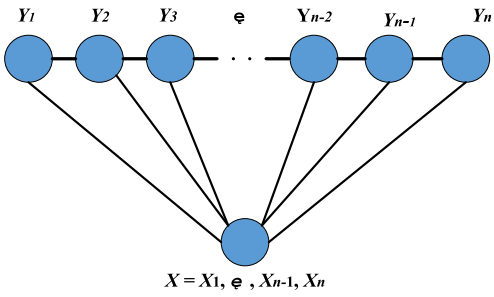


FIGURE 3. Chain structured of CRF.

Given a sequence of observations X, the probability of a particular marker sequence Y can be expressed as:

$$\exp\left[\sum\nolimits_j \beta_j t_j\left(y_{i-1}, y_i, X, i\right) + \sum\nolimits_k \mu_k s_k\left(y_i, X, i\right)\right]$$

where, $t_j(y_{i-1}, y_i, X, i)$ is a transition feature function which indicates the transition probability of the marked sequence of the observed sequence X at the i to i-1 position. $s_k\ (y_i, X, i)$ is a status feature function representing the probability of marking for the position of the observation sequence X whose i is. β_j and μ_k are the weights of t_j and s_k. When defining a feature

function, we define a set of $\{0, 1\}$ binary features $b(X, i)$ about the observed sequence to represent the distribution of features in the training sample:

$$b(X, i) = \begin{cases} 1, & X\text{'s } i \text{ position a specific word} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where i represents a dimension or feature of X.

Therefore, the transition feature function can be expressed as:

$$t_j(y_{i-1}, y_i, X, i) = \begin{cases} b(X, i), & y_{i-1},\ y_i \text{ satisfy conditions} \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

namely, when y_{i-1} and y_i satisfy the transition condition and X_i is the specific word, the transition feature function takes 1; otherwise, it is 0.

For the convenience of description, the status feature function can be written as follows:

$$s(y_i, X, i) = s(y_{i-1}, y_i, X, i) \quad (4)$$

If both the transition feature function and the status feature function are abstracted as f(x), then:

$$F_j(Y, X) = \sum_{i=1}^{n} f_i(y_{i-1}, y_i, X, i) \quad (5)$$

Therefore, the conditional probability of the conditional field is:

$$p(Y|X, \beta) = \frac{1}{z(x)} \exp\left[\beta_j \cdot F_j(Y, X)\right] \quad (6)$$

where $z(x)$ is the normalization factor, and β_i represents the corresponding coefficient.

Based on the above definitions and derivations, the NER extraction can be performed on the device. There are three main steps: 1) a corpus must be prepared, which comes from crawling information about manufacturing equipment from the China Smart Manufacturing website; 2) a program that automatically organizes HTML markup data and removes spaces is designed to improve the missing and erroneous data; 3) The Chinese word segmentation and part-of-speech tagging are completed by the jeba word segmentation tool, and the model is trained using the CRF ++ tool.

C. IDENTIFICATION RELATION

As shown in Fig. 2, this part is mainly focused on the relationship identification between the smart devices. These relationships can be a logical relationship, such as inclusion relationships, causal relationships, category relationships, and premise relationships, which are important for engineers and line operators. The development of a product requires the intensive knowledge involved in the product's lifecycle management, which is often a complex, fuzzy, and iterative process. And the product development process will require further improvement of the knowledge system. The optimal ternary combination is selected as the relationship recognition between the devices by identifying the triples of the entity attribute value, the entity feature, and the composition of the entity object in the sentences and calculating the similarity of the triples.

In the preprocessing stage of the web text, first classify the text, delete the stop words, and use the part of speech. Then, the value was filtered to determine whether the sentence is a numeric type. And finally, the attribute values of the entities are extracted from the text according to the customized Chinese vocabulary. Therefore, we completed the syntactic analysis, including removing the digital noise words, juxtaposed related nouns, and supplementary components in the sentences, and corrected the errors in the part-of-speech tagging and the predicate complement components of the sentence. Obviously, it is important for obtaining a set of sentence fragments to decompose complex sentences into simple sentences according to relevant rules.

On this basis, we propose the following three points for the segment collection of sentences: 1) By extracting the nouns in the sentence, we construct a possible entity object by using a selection tree algorithm, thereby obtaining a candidate set of the entity object; 2) We extract the verb or verb phrase in the sentence, or the nearest noun from the left of the number, which can be used as a noun feature; 3) Finally, a triple is formed, and the similarity between the triples is calculated to identify a set of device relationships.

Chinese sentences are expressed in a variety of ways. For a more prepared extraction relationship, we first make the following definitions:

Definition 1: A sentence containing three or more entities and entity attribute values is called a complex sentence.

Definition 2: A sentence containing two or less entities and attribute values is called a simple sentence.

According to the above definitions, this paper introduces the following extraction rules through corpus analysis, NLP, and Chinese grammar knowledge:

Rule 1: Suppose a Chinese sentence conforms to the ''(device entity, feature, attribute value)'' mode. As shown in Fig. 4, $O(NN)$, $F(NN/VV)$, $N(CD)$ represent entity, feature and attribute value, respectively, and there are no redundant options for the path extracted by the relationship. According to Chinese grammar, except for quantifiers, nouns, and verbs, all other words are deleted, and the remaining parts are sorted according to their position in the original sentence, and the result is extracted.

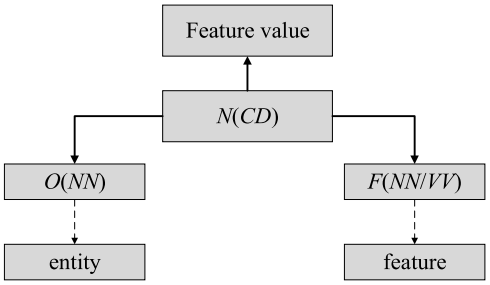


FIGURE 4. The situation of rule 1.

Rule 2: If a sentence in the text meets the ''((entity 1, entity 2, ..., entity i), eigenvalue, attribute value)'' mode, then the prepositional institution in front of the predicate points to the subject. As shown in Fig. 5, $O_1(NN)$, $O_2(NN)$, $O_3(NN)$, $O_4(NN)$ represent entity 1, entity 2, entity 3, entity 4, respectively. And $F(NN/VV)$, $N(CD)$ correspondingly represent feature and attribute value. When using the word segmentation tool for word segmentation, the part before the predicate may be separated into multiple nouns, which may be part of the feature object. At the same time, corresponding to the relationship between multiple entities, multiple entities are in a side-by-side relationship located in front of the feature words. Therefore, the result of extracting the feature object candidate set is selecting several combinations in order from ''(entity 1, entity 2, ..., entity n)'', and the whole entity relationship extraction is ''(entity combination option, feature word, attribute value).''

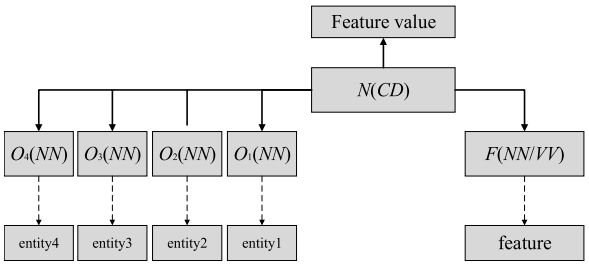


FIGURE 5. The situation of rule 2.

In this part, we mainly use the features of grammar analysis to manually extract the attribute relationships between the manufacturing device entities. For the information text of equipment from the Internet, due to the particularities of Chinese, we use a lot of Chinese word formation and sentence ideas to enrich the relationship between entities as much as possible. In this way, the KG can be more complicated, and it contains more abundant data relationships. At the same time, our information system for KnowIME can better reflect its value.

Syntactic analysis method based on the above analysis of entity relations is more ordinary method, but it is very efficient and simple, we don't need cluster servers to run our code. Compared to the KG based on deep learning, they need more computing resources, such as CPU, GPU and memory, and the model usually is more complicated.

V. CASE STUDY

In this section, we will build a complete KnowIME information system through actual cases, which is mainly used for data storage and visualization through the graph database Neo4j. In addition, we consider the importance of graph associations and use the shortest path algorithm of the graph to recommend and retrieve device information that is most relevant to that information. The basic hardware and software configurations used in the experiment are shown in Table 1.

In the constructed system, the designed nodes include device entity nodes (such as clothing, milling machines) and urban entity nodes. The left subtree of the root node is the intelligent manufacturing equipment, the right subtree is the node of the province, the child nodes of the left subtree represent the device entity to create the child node according to the classification from large to small, and the child of the right subtree node is the urban node. In the previous section, we used syntactic analysis to find the triples (object entities, feature words, and attribute values) that appear in the sentence, and calculated the similarity of these triples to determine the relationship. From the intelligently manufactured news corpus, we used the unsupervised syntax analysis method proposed above to obtain the relevant relationships, as shown in Table 2.

TABLE 1. Experiment environment configuration information.

Component	Configuration
Hadoop cluster	Apache-hadoop-2.7.1
	Apache-Zookeeper-3.4.7
	Apache-Hive-1.2.2-bin
	Apache-Hbase-0.98.17-hadoop2
Spark	Spark-2.3.0
Mysql	Mysql Server-5.5.27
Web application server	apache-tomcat-7.0.91(64-bit)

TABLE 2. The relations of entities.

Entity	Relation	entity
Lathe	paralleling relation	milling
Lathe	assisting relation	robot
Milling	is from	city
production environment	is made of	equipment

Since the corpus data suitable for extracting relationships between device entities is very small, the new data of the device is skewed and only a simple relationship similar to Table 2 can be obtained. In the process of construction, it was also necessary to artificially add some relationship data of device information and complete the device KG information according to the additional information. In addition, we added an extra feature to the system, which used the shortest path algorithm of the graph to make recommendations. Specifically, the forms of the algorithm mainly include determining the shortest path problem of the starting point, determining the shortest path of the end point, determining the shortest path problem of the starting point and the ending point, and the global shortest path problem. Then, through the shortest path algorithm, when the user retrieves the related information of the device, the shortest path ranking is calculated according to the directed graph between the entities, and the first three results are returned.

As shown in Fig. 6, the KG information of the lathe sub-class equipment is shown, and the KG of the milling sub-device is similar. The green circle in the figure represents the sub-class equipment, and the blue circle represents the

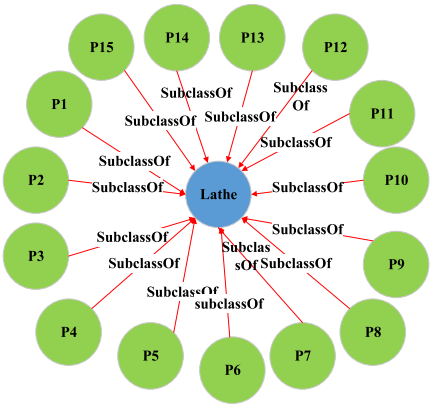


FIGURE 6. Graph structure of lathe subclasses.

parent class. entity1{*Id*: String, *product_Name*: String, *company*: String, *product_Price*: String, *product_place*: String}.

The overall process of device KG retrieval is shown in Fig. 7. The system mainly completed the above processes through query, parse, compile, optimize, compute, bind, and visualization. Specifically, if we enter ''lathe (related parts)'', the system first analyzes the user's input according to the syntax of Neo4j, and then compiles the query program into a file recognizable by the Java virtual machine through the compiler. Then, we needed to optimize the relevant program and input it to Spark's GraphX to calculate the data. And the result of the calculation is stored as a data file in the JavaScript object symbol (json) format. At the same time, Jquery's Ajax asynchronous load data intuitively returns the result for the user. Finally, the system calculates the node information of the shortest path according to the Dijkstra shortest path algorithm, which is calculated according to the idea of the BFS. Specifically, The algorithm starts with a source node and then calculates the distance from its neighbor to it by level; Then take the calculated node as the starting point and calculate the distance from other nodes to the starting point; This continues until it has traversed all nodes and returns the node closest to the source node. In general, the system uses the input node ''lathe'' as the starting node, and then finds the node closest to the node based on the node distance settlement and makes recommendations. (It should be noted that because the attribute name of the node is Chinese, the following figure is abstracted as a character.)

Based on a KG consisting of basic equipment information, the system can provide relevant information services to specific engineers or staff. For example, when engineers need to develop an automated production line for automotive production, they need to know the information about each device and the information about the supplier. At this point, they can use the external query and device association information recommendation service provided by KnowIME to obtain complete information about the device, which can save a lot of development time and improve equipment assembly efficiency. This process is actually information retrieval, and mainly relies on the characteristics of the graph database Neo4j to perform similarity matching and shortest path calculation between graph nodes, which is different from traditional information retrieval. Fig. 7 shows the basic implementation of the entire information index, and the final visualization results just show the information we need to query. When we enter the word ''lathe'', the system encapsulates the query syntax of neo4j and loads it through the Java virtual machine at the bottom.

When using Cypher for information retrieval in Neo4j, the efficiency of data retrieval decreases when the graph structure reaches three levels. Therefore, we consider using graph algorithms to optimize related programs. Meanwhile, distributed computing is used to run the non-relational database Neo4j to improve the storage and calculation of graph nodes. In addition, in the KG information retrieval process, Spark's GraphX is used to accelerate the traversal of the graph, thereby improving the efficiency of KnowIME information retrieval.

And finally, as shown in Fig. 8, we compare KnowIME system with the traditional database-based equipment information retrieval method from the following two dimensions:

(a) Retrieval efficiency: The efficiency of indexing specified device related information from mass data (including operation process and execution process);
(b) Information richness: The number of results finally returned when indexing the specified device related information from the mass data.

As can be seen from the results, on one hand, the related data of KnowIME system is connected by knowledge graph, so the retrieval efficiency of KnowIME is higher than the traditional database-based retrieval method. On the other hand, the KnowIME system knowledge map can effectively organize more related information of the device, so it can obtain more comprehensive information of the device.

VI. DISCUSSION AND OUTLOOK

This paper proposes a KG-based manufacturing equipment information query system, which provides a means for intelligent equipment information acquisition and effective utilization. However, due to the complexity, heterogeneity and changeability of the manufacturing system, the constructed KnowIME system still has certain limitations. On the one hand, the system focuses on the level of stand-alone equipment, and the information obtained is still relatively limited. On the other hand, the information of the manufacturing system is very scattered, and the acquisition of a large amount of valid data still has problems. In this paper, we use web crawler technology to get a lot of data from the Internet, which solves the problem of insufficient data to some extent, but it will undoubtedly reduce the accuracy of the system.

Therefore, in the future, we will continue to study in the following aspects. First of all, the accuracy of device relationship attribute extraction is further improved, and a more fine-grained KG structure is constructed, so that the system can perform more accurate device information recommendation. And then, by improving the quality of equipment information retrieval and reducing the complexity of graph

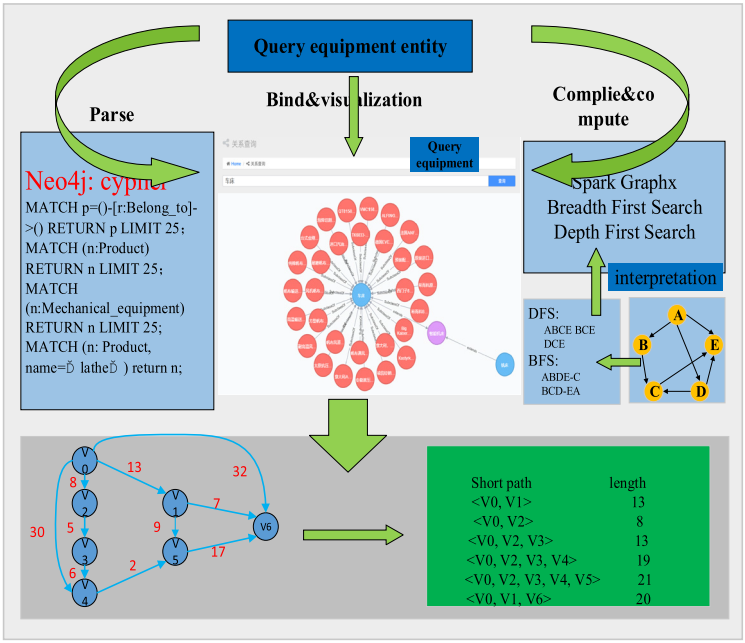


FIGURE 7. The process of querying information.

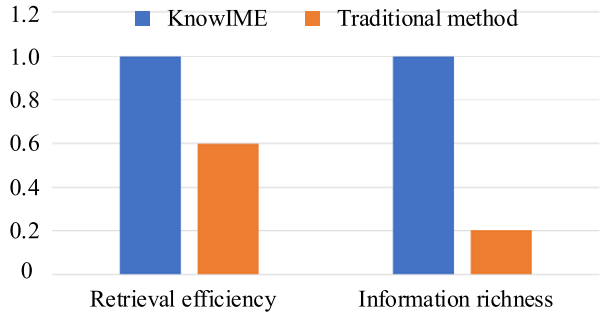


FIGURE 8. The comparison of knowIME and traditional method.

database query, the proposed system is optimized to improve the real-time information query. In addition, we will consider building a KG-based information system at the system level to support the operation and scheduling of the entire smart factory.

VII. CONCLUSION

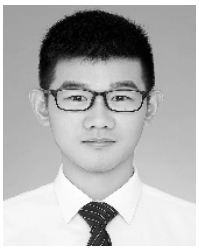
This paper aims to solve the problem of equipment information utilization and management that is faced by intelligent manufacturing related enterprises in the product design and development process. Based on the concepts of KG and graph database tool, this paper proposes to build a device information system for KnowIME. This article first collects a large number of device entities, attribute data and intelligently manufactured news corpus data from the China Smart Manufacturing website. The unstructured text data is then processed by preprocessing means, such as regular expression filtering and normalized text formatting, and integrated into a text corpus. Then, the statistical probability map model algorithm CRF is used to extract the entities of the devices and the cities, and the parsing method is used to calculate the similarity between the entities for relationship extraction. Finally, KG was built with the Neo4j tool and the results of the device information query were visualized. In addition, the associated device information is calculated by the shortest path algorithm and recommended to the user. In this way, the user can obtain the search device related information by querying the input device name, which undoubtedly improves work efficiency and resource utilization.

REFERENCES

[1] B. Chen, J. Wan, L. Shu, P. Li, M. Mukherjee, and B. Yin, "Smart factory of industry 4.0: Key technologies, application case, and challenges," *IEEE Access*, vol. 6, pp. 6505–6519, 2018.
[2] J. Wu, Z. Chen, and M. Zhao, "Information cache management and data transmission algorithm in opportunistic social networks," *Wireless Netw.*, vol. 25, no. 6, pp. 2977–2988, Feb. 2018.
[3] J. Wan, D. Zhang, Y. Sun, K. Lin, C. Zou, and H. Cai, "VCMIA: A novel architecture for integrating vehicular cyber-physical systems and mobile cloud computing," *Mobile Netw. Appl.*, vol. 19, no. 2, pp. 153–160, Feb. 2014.
[4] J. Wan, S. Tang, D. Li, S. Wang, C. Liu, H. Abbas, and A. V. Vasilakos, "A manufacturing big data solution for active preventive maintenance," *IEEE Trans Ind. Informat.*, vol. 13, no. 4, pp. 2039–2047, Aug. 2017.
[5] J. Wan, J. Li, M. Imran, D. Li, and Fazal-E-Amin, "A blockchain-based solution for enhancing security and privacy in smart factory," *IEEE Trans Ind. Informat.*, vol. 15, no. 6, pp. 3652–3660, Jun. 2019.
[6] J. Wan, B. Chen, M. Imran, F. Tao, D. Li, C. Liu, and S. Ahmad, "Toward dynamic resources management for IoT-based manufacturing," *IEEE Commun. Mag.*, vol. 56, no. 2, pp. 52–59, Feb. 2018.
[7] J. Wu, G. Yu, and P. Guan, "Interest characteristic probability predicted method in social opportunistic networks," *IEEE Access*, vol. 7, pp. 59002–59012, 2019.
[8] P. Guan and J. Wu, "Effective data communication based on social community in social opportunistic networks," *IEEE Access*, vol. 7, pp. 12405–12414, 2019.
[9] J. Wu and Z. Chen, "Sensor communication area and node extend routing algorithm in opportunistic networks," *Peer-Peer Netw. Appl.*, vol. 11, no. 1, pp. 90–100, Oct. 2016.
[10] M. Xia, T. Li, Y. Zhang, and C. W. de Silva, "Closed-loop design evolution of engineering system using condition monitoring through Internet of Things and cloud computing," *Comput. Netw.*, vol. 101, pp. 5–18, Jun. 2016.

[11] M. Xia, T. Li, L. Xu, L. Liu, and C. W. de Silva, "Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 101–110, Feb. 2018.
[12] Y. Hassanzadeh-Nazarabadi, A. Küpyü, and Ö. Özkasap, "Decentralized and locality aware replication method for DHT-based P2P storage systems," *Future Gener. Comput. Syst.*, vol. 84, pp. 32–46, Jul. 2018.
[13] J. Wu, Z. Chen, and M. Zhao, "Weight distribution and community reconstitution based on communities communications in social opportunistic networks," *Peer-Peer Netw. Appl.*, vol. 12, no. 1, pp. 158–166, Apr. 2018.
[14] J. Qi, A. Liu, and Y. Lei, "Research on XML schema-based manufacturing information integration specification," *Comput. Integr. Manuf. Syst.*, vol. 11, no. 4, pp. 565–571, 2005.
[15] Q. Qi and F. Tao, "Digital twin and big data towards smart manufacturing and industry 4.0: 360 degree comparison," *IEEE Access*, vol. 6, pp. 3585–3593, 2018.
[16] L. Qiao, L. Yang, D. Hong, L. Yao, and Q. Zhiguang, "Knowledge graph construction techniques," *J. Comput. Res. Develop.*, vol. 53, no. 3, pp. 582–600, 2016.
[17] R. Jain, S. Iyengar, and A. Arora, "Overview of popular graph databases," in *Proc. 4th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, Jul. 2013, pp. 1–6.
[18] A. Singhal, "Introducing the knowledge graph: Things, not strings," Off. Google Blog, CA, USA, Tech. Rep. 16, 2012.
[19] P. Chen, Y. Lu, V. W. Zheng, X. Chen, and B. Yang, "KnowEdu: A system to construct knowledge graph for education," *IEEE Access*, vol. 6, pp. 31553–31563, 2018.
[20] J. Wan, B. Yin, D. Li, A. Celesti, F. Tao, and Q. Hua, "An ontology-based resource reconfiguration method for manufacturing cyber-physical systems," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 6, pp. 2537–2546, Dec. 2018.
[21] (2017). *IBM Waston*. [Online]. Available: https://www.ibm.com/ watson/
[22] Z. J. Zhang, "Graph databases for knowledge management," *IT Prof.*, vol. 19, no. 6, pp. 26–32, Nov. 2017.
[23] H. Kim, "Towards a sales assistant using a product knowledge graph," *Web Semantics, Sci., Services Agents World Wide Web*, vols. 46–47, pp. 14–19, Oct. 2017.
[24] Z. Wu, J. Liao, W. Song, H. Mao, Z. Huang, X. Li, and H. Mao, "Semantic hyper-graph-based knowledge representation architecture for complex product development," *Comput. Ind.*, vol. 100, pp. 43–56, Sep. 2018.
[25] Z. Liu, C. Xiong, M. Sun, and Z. Liu, "Entity-duet neural ranking: Understanding the role of knowledge graph semantics in neural information retrieval," 2018, *arXiv:1805.07591*. [Online]. Available: http://arxiv.org/abs/1805.07591
[26] A. Han, D. Wong, and L. Chao, "Chinese named entity recognition with conditional random fields in the light of chinese characteristics," in *Language Processing and Intelligent Information Systems*. Springer, 2013, pp. 57–68.
[27] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*. [Online]. Available: http://arxiv.org/abs/1412.3555
[28] J. Gao, M. Li, C.-N. Huang, and A. Wu, "Chinese word segmentation and named entity recognition: A pragmatic approach," *Comput. Linguistics*, vol. 31, no. 4, pp. 531–574, Dec. 2005.
[29] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF models for sequence tagging," 2015, *arXiv:1508.01991*. [Online]. Available: http://arxiv.org/abs/1508.01991
[30] C. Quirk and H. Poon, "Distant supervision for relation extraction beyond the sentence boundary," 2016, *arXiv:1609.04873*. [Online]. Available: http://arxiv.org/abs/1609.04873
[31] D. Zeng, K. Liu, Y. Chen, and J. Zhao, "Distant supervision for relation extraction via piecewise convolutional neural networks," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 1753–1762.
[32] J. Wan, S. Tang, Q. Hua, D. Li, C. Liu, and J. Lloret, "Context-aware cloud robotics for material handling in cognitive industrial Internet of Things," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2272–2281, Aug. 2018.
[33] J. Wan, B. Chen, S. Wang, M. Xia, D. Li, and C. Liu, "Fog computing for energy-aware load balancing and scheduling in smart factory," *IEEE Trans Ind. Informat.*, vol. 14, no. 10, pp. 4548–4556, Oct. 2018.
[34] S. Beitzel, E. Jensen, and D. Grossman, "Retrieving OCR text: A survey of current approaches," in *Proc. Symp. Document Image Understand. Technol.*, 2003, pp. 1–6.
[35] Y. Adi, E. Kermany, Y. Belinkov, O. Lavi, and Y. Goldberg, "Analysis of sentence embedding models using prediction tasks in natural language processing," *IBM J. Res. Develop.*, vol. 61, no. 4/5, pp. 3:1–3:9, Jul. 2017.
[36] S. Wu, M. Liu, and H. Hu, "Unsupervised extraction of attribute-value entity relation from chinese texts," *Wuhan Univ.*, vol. 62, no. 6, pp. 552–560, 2016.
[37] R. Grishman, "The NYU system for MUC-6 or where's the syntax?" *Proc. 6th Conf. Message Understand. Assoc. Comput. Linguistics*, 1995, pp. 167–175.
[38] S. Brin, "Extracting patterns and relations from the world wide Web," in *Proc. Int. Workshop World Wide Web Databases*. Berlin, Germany: Springer, 1998, pp. 172–183.
[39] J. Xu and J. Zhu, "Astronatuics named entity recognition based on CRF algorithm," *Electron. Des. Eng.*, vol. 25, no. 20, pp. 42–46, 2017.
[40] W. Jin and H. H. Ho, "A novel lexicalized HMM-based learning framework for Web opinion mining," in *Proc. 26th Annu. Int. Conf. Mach. Learn. (ICML)*, 2009, pp. 465–472.
[41] L. Xie, M. Zhou, and M. Sun, "Sentiment analysis of chinese micro Blog and its features extraction," *J. Chin. Inf. Process.*, vol. 26, no. 1, pp. 73–84, 2012.
[42] Y. Zhou, Y. Guo, and X. Huang, "Chinese and english based NP recognition based on a maximum entropy model," *J. Comput. Develop.*, vol. 40, no. 3, pp. 440–446, 2003.
[43] Y. He, C. Luo, and B. Hu, "Geographic entity recognition method based on CRF model and rules combination," *Comput. Appl. Softw.*, vol. 32, no. 1, pp. 179–185, 2015.

HEHUA YAN has directed three research projects, including the Natural Science Foundation of Guangdong Province. She is currently an Associate Professor with the School of Electrical Technology, Guangdong Mechanical and Electrical Polytechnic, China. Thus far, she has authored or coauthored more than 20 scientific articles. Her research interests include embedded systems, the Internet of Things, and cyber-physical systems.



JUN YANG received the B.A. degree in mechanical engineering from the Wuhan University of Technology, China, in 2017. He is currently pursuing the M.S. degree with the School of Mechanical and Automotive Engineering, South China University of Technology, China. His research interests include smart factory, cyber-physical systems, the industrial Internet of Things, and industrial big data.

JIAFU WAN has directed 20 research projects, including the National Key Research and Development Program of China, the Key Program of National Natural Science Foundation of China, and the Guangdong Province Key Areas Research and Development Program. He is currently a Professor with the School of Mechanical and Automotive Engineering, South China University of Technology, China. Thus far, he has published more than 160 scientific articles, including more than 100 SCI-indexed articles, more than the 40 IEEE TRANSACTIONS/JOURNAL ARTICLES, 20 ESI Highly Cited Articles, and four ESI Hot Articles. According to Google Scholar, his published work has been cited more than 10 000 times. His research interests include cyber-physical systems, intelligent manufacturing, big data analytics, industry 4.0, smart factory, and cloud robotics. His SCI other citations (sum of times cited without self-citations) reached 2500 times according to Web of Science Core Collection. He is also listed as a Clarivate Analytics Highly Cited Researcher, in 2019. He is also an Associate Editor of the IEEE/ASME TRANSACTIONS ON MECHATRONICS, the *Journal of Intelligent Manufacturing*, and *Computers & Electrical Engineering*, and an Editorial Board of *Computer Integrated Manufacturing Systems*.