Research on Data Link Knowledge Graph Technology and Its Application Based on Data

Lidong Zhang*, Dan Wang[a]

Institute of Systems Engineering Academy of Military Science (AMS), Beijing, China

*xxddxdd@yeah.net; [a]wdn_719@126.com

Abstract. This paper first analyzes the dilemma of digital construction of data link equipment under the current digital background of our army and the necessity of studying the knowledge graph of data link. Then it introduces the current research status and development process of knowledge graph technology at home and abroad, combs the definition, construction, storage and application of knowledge graph, and the update and maintenance of the whole life cycle, and expounds the relationship between knowledge graph and ontology. On this basis, combined with the characteristics of data link equipment, this paper puts forward the construction method of data link knowledge graph based on data, which is of great significance for the later digital construction of our army's data link equipment, improving the effect of data governance, improving the efficiency of equipment data use, and assisting the decision-making of relevant data link practitioners.

Keywords: data link; Knowledge graph; Data governance

1 INTRODUCTION

With the rapid development of science and technology, modern war has changed greatly in its ways and methods compared with traditional war. Modern war has evolved into information centric war. The use of a large number of advanced military equipment has further expanded the war space, and military operations can spread all over the land, sea and even space. In this context, the importance of data link equipment in modern war is self-evident. There are many types of data link equipment. The deep hinge with the platform and complex application scenarios make the management and development of data link equipment full of difficulties. Therefore, in the context of the current military digitization and modernization transformation, it is extremely urgent to find digital and visual management methods for relevant data link equipment. Knowledge graph technology can handle complex data management problems with strong semantic processing ability, and its construction results have user-friendly visualization ability [1].

Since the "14th five year plan", China's data link equipment construction has made considerable progress, and digital construction has become the development direction of China's data link equipment and organization. Data link equipment is the core of

A. Haldorai et al. (eds.), *Proceedings of the 2024 3rd International Conference on Artificial Intelligence, Internet and Digital Economy (ICAID 2024)*, Atlantis Highlights in Intelligent Systems 11,
https://doi.org/10.2991/978-94-6463-490-7_57

modern information war. It is deeply intertwined with reconnaissance and early warning system, command and control system and main war weapons and equipment. Through the digital construction of data link equipment, it can quickly drive multi domain joint operations, and has gradually become the core engine to win the future intelligent war. With the great attention and investment of the state, the digital construction of data link equipment and its related technologies have received extensive attention and attention from many domestic military experts and researchers. However, due to the asynchronous development of data link equipment in China, the large span of development, the large difference of installed platforms, the decentralized organization and management and other factors, there are complex data collection environment, diverse types, high correlation, strict security and confidentiality requirements in the short term The characteristics of strong real-time antagonism lead to unclear data base, low data quality, difficulty in efficient integration, difficulty in data distribution, difficulty in accurate analysis, and insignificant data value-added effect. Complex high-dimensional data are difficult to be directly and visually used by commanders, and the data output can not fully meet the requirements of digital and information warfare, which has brought severe challenges to the digital construction of data link equipment, It seriously restricts the innovation ability of data link equipment in the digital transformation environment.

The data link equipment management under the background of digital transformation centers on the data link equipment entity, and takes its relevant industrial departments, design and R&D personnel, relevant user organizations, and daily data maintenance as management elements. The data link and its management elements have the characteristics of large amount of data, many related parties, complex interrelationships, and complex types. The relevant management object of data link equipment is the whole life cycle of data link equipment from R&D to scrapping, covering the data link and the integrated application platform and related systems with data link hinges, including all factor related parties and related party transaction including production and use. However, there are many types and a large number of users related to the R&D, production and use of data link equipment, including government users, industrial departments and user forces. Knowledge graph technology is a new knowledge management technology based on graph structure, which has the characteristics of clear relationship, complex processing objects and high visualization. Studying and promoting the construction of data link knowledge graph based on data is conducive to promoting the formation of a more efficient and clear information management system for data link equipment, so as to effectively improve the drawbacks of existing data link management. Effectively improve the governance ability of data link data to assist decision-making.

2 RESEARCH STATUS OF MILITARY KNOWLEDGE GRAPH

The research results of related knowledge graph in the domestic military field are limited and develop slowly. In 2015, Feng Yun[2-3] and others proposed a knowledge graph

construction method for semi supervised learning entity recognition based on conditional random fields. This method uses military related terms and expression characteristics as a priori to establish relevant feature sets, and uses priori sub points and priori rules to correct the recognition results, so as to achieve the purpose of military knowledge graph construction. Based on the existing open source military equipment database, Ding and Zhao[4] put forward a brand-new military related equipment knowledge graph construction model in 2018. However, based on the existing open source data characteristics, although the organizational structure of this model has some reference value for the subsequent problem solving of military knowledge graph, its traditional syntactic dependency analysis method is difficult to adapt to the current large-scale military text application background. In 2019, Li[5] and others proposed a new method for constructing military knowledge graph by using the two-way long and short-term neural network model based on attention mechanism, combining vector words with word vectors on the corpus, and achieved 87% of the relevant accuracy in the test stage. In 2020, Yin xuezhen, Zhaohui, Zhao junbao[6] and others proposed a method of constructing military knowledge graph based on multi neural network collaboration. This method considered the prior knowledge of relevant experts, and built military corpus based on microblog data based on military corpus set, combined with domain expert knowledge, and combined with various neural networks such as Bert and bilstm for entity recognition, It has achieved high test accuracy on different data sets.

By integrating the above research results, it is not difficult to find that the research on the military knowledge graph lags behind the research in other fields of the knowledge graph. At the same time, there is a lack of expertise graph research in the field of data link. The direct transplantation of relevant technologies is difficult to fully mine the relevant value of data in the field of data link, and there is a lack of systematic data link equipment management index graph research. Based on the characteristics of data link equipment, this paper proposes a scheme for the construction of data link domain knowledge graph, and explores its value[7].

3 OVERVIEW OF KNOWLEDGE GRAPH TECHNOLOGY

The essence of knowledge graph technology is user-friendly visual presentation of structured database. It strives to show real concepts through graphical means and ways, and clearly present the relationship of entities through connections. For users, knowledge graph technology effectively conforms to the individual's cognition of the real world, effectively and reasonably presents the relationship of massive knowledge and completes the management of related entities. The basic network structure of the knowledge graph is shown in the figure 1. In addition, the network storage structure of knowledge graph makes it scalable, and it is very convenient to modify or add new knowledge entities [7].

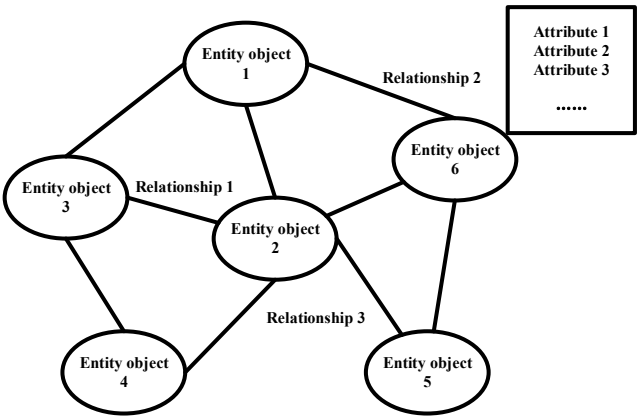

Fig. 1. network structure of knowledge graph

The basic combination of the knowledge graph is a triple of [Entity object 1— Relationship — Entity object 2] and [Entity object—Attribute name—Attribute value]. As shown in the above figure, entity objects are the most basic elements of the knowledge graph. Each node in the knowledge graph network represents a different entity object, which corresponds to a classification or concept individual in the real world, Each entity element has a unique ID to distinguish it from other entities. The connection between different entity elements in the network structure is abstracted as the relationship between different entity elements, and the entity elements form a complete and clear structured knowledge network through such links. Each entity element can protect different attributes. Attributes describe the internal characteristics of different entity elements. For data link equipment, they may be data link type, bandwidth, delay, transmission distance, etc.

The process of establishing a knowledge graph is to extract structured and unstructured target data (including text, tables, web pages, etc.) through different technical processing, store the useful information contained in it in the form of triples of the knowledge graph, and analyze its correlation, so as to form the relevant knowledge system corresponding to the analysis target, It is convenient for the further management of relevant personnel and the secondary utilization of knowledge. The technical process of its construction is shown in Figure 2. Knowledge graph constructs massive data sources related to the construction object. Firstly, it carries out information sobbing on them, extracts the key knowledge of entity objects, and then analyzes their relationships. In the process, it completes the knowledge fusion related to knowledge aggregation and alignment disambiguation at the same time. Finally, it uses the formed graphical database to complete the construction and storage of knowledge graph. Finally, with the continuous updating of domain knowledge and the further expansion of research objects, the construction of knowledge graph also needs to consider the related issues of adding new knowledge, and constantly add brand-new knowledge to the

knowledge graph that has been completed, so as to continuously improve, update and maintain the knowledge graph formed in this field[8].

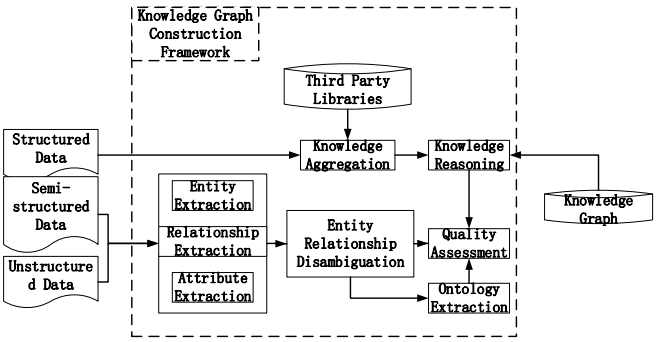

Fig. 2. flow chart of knowledge graph

4 DATA LINK EQUIPMENT KNOWLEDGE GRAPH

Data link is the core equipment of information centric warfare in modern war and the decisive weapon to win the future intelligent war. The data link equipment has complex data acquisition environment, diverse types, high correlation, strict security and confidentiality requirements, and strong real-time antagonism. In order to solve the problems caused by these characteristics, such as unclear data base, difficult accurate analysis, difficult security protection, lack of effective design and management, and insignificant data value-added effect, and ensure that complex high-dimensional data can also be used by relevant projects and commanders to meet the requirements of digitization Based on the above requirements, this study explores the method of constructing data link knowledge graph based on data on the basis of the existing achievements in the early construction of data management related to technology governance[9].

4.1 Research on the construction of data link knowledge graph based on data

Transaction data can be divided into metadata, internal data and external data, structured data and unstructured data by selecting different classification dimensions according to data characteristics.

According to the sovereign attribute of data, it can be divided into external data and internal data. The external data is the data that is not generated and processed by the data link system but needs reference and reference. The external data exists objectively, and its generation and modification are not affected by the transaction activities in the data link field, such as foreign military information, national defense strategy, equipment regulations, policies and regulations, as well as from the design, development External information such as data collected during operation. Internal data is the data generated and processed by various transaction activities within the data link system. It

is the data generated and defined in the whole life cycle of the construction and application of the data link system. Internal data is the focus of transaction data governance[10].

According to the storage property of data, it can be divided into structured data and unstructured data. Structured data is the data that can be expressed and realized by using two-dimensional table structure, which is usually stored in relational databases. Unstructured data is the data that is difficult to be expressed by using two-dimensional table structure of the database, and the form is relatively unfixed, including various heterogeneous format files such as unformatted text, various format documents, images, audio, video, etc, Compared with structured data, it is more difficult to standardize and understand, so intelligent information technology is required to match it in storage, retrieval and consumption. Unstructured data in the data link field includes documents (various reports, plans, drawings, design files, audio and video multimedia materials, etc.), models (executable files, library files, configuration files, algorithm files, etc.), pictures, audio, video, etc.

The structured data of the data link system can be divided into basic data, master data, transaction data, report data, monitoring data and rule data. The specific data classification structure is shown in Figure 3.

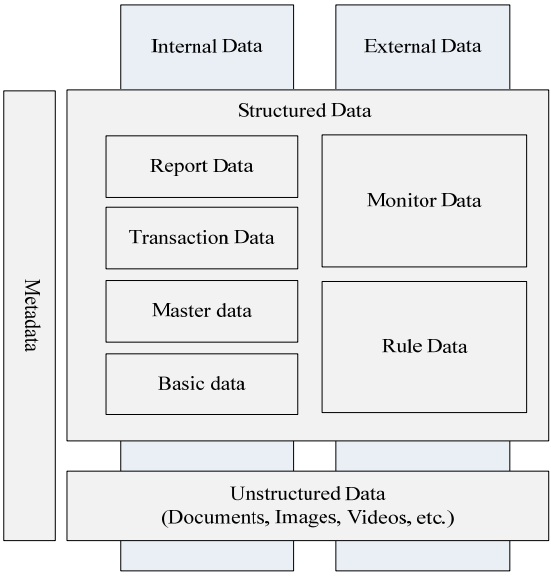

Fig. 3. data link technology governance data classification architecture

The knowledge graph in the field of data link can be constructed in two ways: top-down and bottom-up. The data base for its construction is the accumulation of prior knowledge about data link equipment in the data management software of the technology governance system. After the preliminary construction, the data link knowledge graph needs to be updated and iterated continuously according to the entity and attribute changes of the data link equipment in the field. Refer to the construction process of the

data link equipment knowledge graph, When the data link domain knowledge graph is iterated, the relevant information is extracted first, that is, the entity, attribute change, relationship change, etc. of the object are extracted from different types of data link equipment, and on this basis, the ternary knowledge expression is formed; Then, the newly formed entity knowledge is integrated and knowledge fusion is completed. In the process of knowledge fusion, in addition to adding the existing graph, alignment and disambiguation are also required, such as identifying and removing different expressions of unified knowledge or attributes[11]; Finally, we need to process the new knowledge, that is, add the quality review of the object, so as to ensure the quality of the entire database and knowledge graph. The technical architecture of data link knowledge graph based on data is shown in Figure 4.

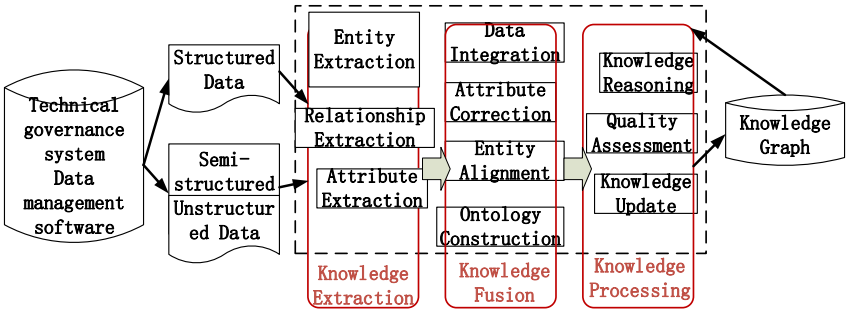

Fig. 4. technical architecture of data link knowledge graph based on data

4.2 Construction of data link knowledge graph based on structured data

The construction of data link knowledge graph based on data takes the structured prior data knowledge of data link as the starting point, and gradually expands on this basis[12]. The design and development of knowledge graph are based on the relevant data link system metadata collected by the data management software in the early stage of technical governance, unstructured reports of master data related projects, and military research reports at home and abroad.

Combined with the characteristics summarized in the technology governance software formed in the early stage, the ontology of knowledge graph is constructed in the way of RDF triples, namely [data link entity 1—Relationship Data—Link Entity 2] and [Data link entity—Data link entity attribute—Data link entity attribute value], with data link system construction projects, data link construction capacity requirements, index construction requirements, etc. as entities, through the relevant data link construction projects with unique ID Data link equipment type, capability item, capability item description and other information, establish the association between entities, and divide multiple attribute values in an entity to represent the entity. At this stage, the design of the relevant mode layer is based on the data management software of the technology governance system, and it will continue to be improved iteratively with the construction changes in the field of data link in the later stage. Taking the knowledge graph of relevant equipment projects as an example, the entity attribute settings and the relationship links

between entities are shown in Table 1. The data layer of the data link knowledge graph is based on the "five databases" and other data.

Table 1. design of data link knowledge graph mode layer

Entity	Attribute	Relationship
Data link system construction	Project name	
	Capability item	
	Indicator 1	
	……	
	The project serial number	Connects the data link related construction projects with the data link capacity-building demand data through the uniquely identified project serial number
Capacity requirements for data link construction	The project serial number	
	Capability item name	
	Capability item description	
	Competency item type	
	Index item number	Connect the data link related index construction requirements with the data link capacity building demand data through the uniquely identified project serial number
Index construction requirements	Index item number	
	Indicator item description	
	……	

4.3 Construction of data link knowledge graph based on unstructured data

The relevant unstructured data in the field of data link includes text information including project reports, plans, and plans, as well as some pictures and video information, among which text information is the main source of unstructured information for constructing the knowledge graph of data link. The main methods to deal with the unstructured information in the field of data link are entity extraction and relationship extraction. Entity extraction is to automatically extract the names, attributes, and attribute values of data link entities from such unstructured data under the constraints of the prior knowledge system in the field of data link. The construction of unstructured data in the data link knowledge graph based on data is mainly aimed at the typical needs of the main businesses in the process of data link construction, as well as the text content of data link related military intelligence, project progress, relevant argumentation materials, the latest military progress at home and abroad, and the report summary documents in the field of data link. Using the construction methods of unstructured data knowledge graph in other fields for reference, the entity link is established by using the algorithm based on semantic similarity, and the unstructured information in the field of data link,

such as typical businesses, project construction management, equipment defect management, data link equipment organization and application, and military intelligence in the field of data link, is included in the knowledge graph representation.

Entity linking method, that is, facing a designated entity X and its possible candidate linked entities e=e1, e2,..., en, the entity linking method first calculates the correlation score between the designated entity and the target entity, then compares the score with the prior threshold value, and associates the entity that meets the threshold requirements with its object, that is, let score (e, x) be the association score of two entities, and if score (e, x) meets the threshold requirements, establish an entity link between e and x.

In this paper, the idea of entity link for data link equipment is as follows. First, the unstructured pen of the data link is processed by word segmentation, and the tag of the target unstructured data is extracted. Then, the semantic similarity is calculated by using the vector space model, which is used to score the correlation and consistency between the two data link entities (e, x). FTF is used to represent the word frequency, that is, the frequency of the entry (keyword) in the text, and FIDF is used to represent the reverse file frequency, which represents the ability to distinguish the categories of entries. The product of FTF and FIDF is used as the value measurement of the feature space coordinate system. The calculation method of TF-IDF is as follows [8]: n_{ij} is the number of occurrences of data represented by the word t_i, | D | is the data overview in the data set, | Di | is the total number of data containing the word Ti in the data set, and w_{ij} is the weight of the word t_i in the data d_j. When calculating the weight of different words in this method, the higher the frequency of the object words in a certain target data and the lower the frequency of the object words in other data, it indicates that the stronger the relevance of the word and the target data, that is, it will give a higher weight to this word. Sort the weight values of all words, so as to select one or more words whose weight value is greater than a certain threshold as the label of the data. This method only takes frequency as the only consideration, and ignores the domain of vocabulary. For example, some words may appear less frequently, but because they have very obvious characteristics of the data link equipment business field, such as "bandwidth", "time slot", "waveform", etc. Such words shall also be regarded as labels of the data link entity. Based on the data dictionary formed by the data management software of the previous technology management system, and integrating the high-frequency words related to the data link business in the field to form a keyword dictionary, the data collection such as the typical question list is trained and studied, and a data classification model is proposed. According to the content and weight of the keyword dictionary, the organizational structure and node behavior of Bayesian network are defined. Then, with the help of the established Bayesian network, a large number of standardized typical problems and other data are trained and learned, so as to achieve label extraction quickly and effectively.

In the process of entity linking, taking the existing data link equipment project construction and allocation information as the entity reference item, taking the typical question list and other data as the target entity, and using the idea of the above label extraction method, the entity reference item and the label vectors Di and DJ of the target entity are formed, i.e., Di={(t1, w1), (t2, w2),..., (tn, wn)}, where n represents the number of labels after the word segmentation of the question list di (i.e., labels with weight values

greater than a certain threshold), tn refers to the word belonging to the question list Di, wn refers to the TF-IDF weight value of the word, and Dj is the same.

If there are k same words in the label vector of Di and Dj, the k words can form two vectors respectively, namely Si={(S1, W1), (S2, W2),..., (Sk, Wk)}, Sj={(S1, W1), (S2, W2),..., (Sk, Ek)}. Then, for Si and Sj, the similarity distance between two keywords is calculated according to the cosine similarity algorithm. The formula is:

The similarity distance of the label vector is used as the entity association score score (e, x) to compare with the preset Association threshold. If it is greater than the preset Association threshold, the entity link between e and x is established. Using prior knowledge to continuously improve the preset threshold, so as to ensure the accuracy and adequacy of entity links to achieve a more ideal effect.

4.4 Visual display of data link knowledge graph

The visual display of data link knowledge graph is the most important application after the construction of knowledge graph, which can provide users with a means to quickly obtain the knowledge of the target data link. At the same time, the secondary development interface is reserved.

Cypher language is used as the data query language. The b/s architecture is used to design relevant application systems, the back end uses the flask framework to build web services, and the front-end visual display technology is implemented by echarts and HTML5.

The purpose of entity search is to match the data link keywords entered by the user of the data link knowledge graph with the data link entities or data link attributes in the knowledge base, find the entities that match the keywords and their categories, and display all other entities related to the target entity, so as to show the complete knowledge of the target data link entities to users.

When a user searches with a data link equipment model or item name as a keyword, it will display its relevant knowledge network centered on the matched data link entity.

4.5 Quality evaluation of data link knowledge graph

In the process of evaluating the quality of data link knowledge graph, users' needs and user experience feedback should be taken as the starting point and foothold of the quality evaluation of knowledge graph, and the objectivity, comprehensiveness and efficiency of the process should be ensured. Among them, users of the Military Commission, users of military services, data link developers of industrial departments, knowledge graph designers, and other interested parties should be included in the evaluation process, taking into account their use experience and needs. Figure 5 shows the quality evaluation process of data link knowledge graph.

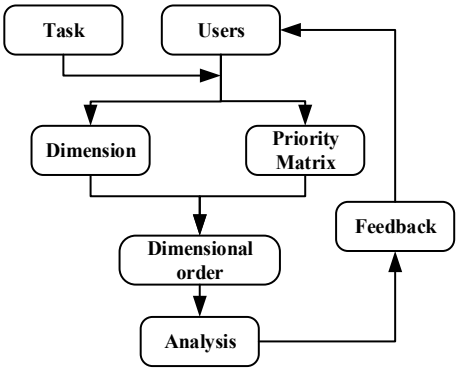

Fig. 5. data link knowledge graph quality evaluation process

The quality evaluation of data link knowledge graph is mainly divided into the following four steps:

1) Collect users feedback and needs during use;

2) Determine the evaluation dimension and priority of the data link knowledge graph;

3) Carry out quality evaluation on the data link knowledge graph according to the evaluation dimension and priority;

4) The result feedback of data link knowledge graph quality evaluation is used to guide the secondary construction of knowledge graph.

5 SUMMARY

Knowledge graph is highly integrated and integrated with a variety of new technologies, including but not limited to natural language processing, expert prior, and learning. Knowledge graph provides a new method for knowledge representation, storage and management. Promoting the application of knowledge graph technology in the field of data link can effectively improve the management ability of data link equipment, assist in the digital transformation and modern governance construction of data link equipment management, further promote the development of data link governance business, and assist in the upper level decision-making related to the field of data link.

REFERENCES

1. Chunxia Zhang Gao Yang Design requirements for information integration of aircraft cooperative ad hoc network data link [J] Aerospace Standardization, 2021, No.183 (01): 1-5
2. Yiming Wei, Tielin Zhu, Lixin Niu Design of a Time Division Ad hoc Network UAV Data Link [J] Electronic World, 2021, No.611 (05): 172-173
3. Furong Yu, Yan Zhang, Xue Jiang, Jinna Qiu Development Trend and

4. Huang Hengqi, Yu Juan, Liao Xiao, et al. Review of Knowledge Graph Research [J]. Computer System Applications, 2019,28 (6): 1-12
5. Zhang Yupeng Research on the Construction Technology of Military Equipment Knowledge Graph [D] Shaanxi: Xi'an University of Electronic Science and Technology, 2021
6. Feng Yuntian, Zhang Hongjun, Hao Wenning, Chen Gang Named Entity Recognition Based on Deep Belief Networks [J] Computer Science, 2021, v.43 (04): 224-230
7. Ding Junyi, Zhao Qingsong, Xia Boyuan, Zou Zhigang Research on the Construction Method of Weapon Equipment Knowledge Graph Based on Open Source Data [J] Command, Control and Simulation, 2020, v.40; No.272 (02): 22-26
8. Li Jianlong, Wang Panqing, Han Qiyu Military Named Entity Recognition Based on Bidirectional LSTM [J] Computer Engineering and Science, 2019, v.41; No.292 (04): 713-718
9. Hyunjin C, Youdan K, Hyounjin K. Genetic Algorithm Based Decentralized Task Assignment for Multiple Unmanned Aerial Vehicles in Dynamic Environments [J]. International Journal of Aeronautical and Space Science, 2021, 12(2):163-174
10. Che Jinli, Tang Liwei, Deng Shijie, etc Construction and Application of Military Equipment Knowledge Graph Based on Encyclopedia Knowledge [J]. Journal of Weapon Equipment Engineering, 2019,40 (1): 148-153
11. G. S. Schrecke, S. Davidson, M. A. Kahn, M.-C.Wang, M.W. Henry, Ip address translation for tactical networks, uS Patent App. 10/063,516 (Aug. 28 2021).
12. I. Eyal, E. G. Sirer, Majority is not enough: Bitcoin mining is vulnerable, Communications of the ACM 61 (7) (2020) 95–102.