A knowledge Graph Construction Approach for Ship Overall Performance Domain

Kangkang Dou[1], Dongsheng Wang[2*], Kun Tang [2], Wei Chen [1]

Kangkang Dou: douk2008@163.com, Dongsheng Wang: jsjxy_wds@just.edu.cn,
Kun Tang: jsj_tk@stu.just.edu.cn, Wei Chen: 18036675917@163.com
[*]Corresponding author's e-mail: jsjxy_wds@just.edu.cn

[1] Jiangsu Automation Research Institute, Lianyungang, Jiangsu, 222061, China
[2] Department of Computer, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu, 212000, China

Abstract. The Ship Overall Performance (SOP) generally refers to the performance that has a decisive effect on the overall indicators of the ship. Knowledge graph is used to describe entities and concepts in the real world and the relationships between them. By constructing a knowledge graph in the SOP domain, it can better meet the business analysis needs of scientific research or engineering personnel in terms of knowledge question-answering and decision-making assistance. Firstly, the data in the SOP domain is analysed, and an ontology construction method based on the seven-step method is proposed, and the ontology construction is completed with the help of protégé tools. Then, knowledge extraction and fusion are carried out on the massive domain text, and finally the RDF triples are mapped to the graph database, and the graph data is stored in the database to realize knowledge visualization, and realize the construction of the knowledge graph in the SOP domain.

Keywords: Knowledge graph, Ontology, Knowledge fusion, Ship overall performance domain

1 Introduction

After entering the 21st century, China has gradually strengthened the construction of a marine power, and the shipbuilding industry has entered a stage of rapid development, which has improved to a certain extent in terms of quality and performance, but compared with the current developed countries, the overall performance is still relatively backward. The overall performance of a ship refers to some indicators related to the overall performance of the ship [1], which play a decisive role, including principal scale, navigation performance, displacement and so on. It plays an important role in the navigation of the ship, and determines the basic performance of the ship to a great extent. At present, knowledge graph [2] is the most effective means to manage knowledge, and the use of knowledge graph can manage the data more conveniently.

In order to meet the business analysis needs of scientific research or engineering personnel in the SOP domain in knowledge question-answering and auxiliary decision-making, and to improve the efficiency of scientific research and the convenience and accuracy of knowledge acquisition, it is of great practical significance to construct the knowledge graph in the SOP

PMBDA 2023, December 15-17, Nanjing, People's Republic of China

DOI 10.4108/eai.15-12-2023.2345390

domain. At present, there are still some difficulties in the process of constructing the knowledge graph in the SOP domain, which are shown in the following points. First, without the existing SOP ontology, the knowledge graph cannot be constructed on the basis of the ontology. Second, because the overall performance of the ship contains a variety of data, and the data source is heterogeneous, there is a situation that one entity corresponds to multiple pronouns and one attribute corresponds to multiple attribute values.

In view of the difficulties existing in the process of constructing the SOP knowledge graph, this paper studies the construction method and related technology of the SOP knowledge graph. Combined with the characteristics of the data in the SOP domain, the information is extracted from the marine data to construct the knowledge graph in the SOP domain.

2 Related work

In recent years, with the explosive growth of information data, the management of knowledge and information has been greatly updated. In 2012, google first put forward a concept of knowledge graph [3], which is used to integrate and summarize a large number of data and knowledge information. Knowledge graph is essentially a concept of semantic network, and it is a knowledge base that can well express entities and concepts in the real world and their relationship. The knowledge graph is a huge semantic network stored in the way of a graph. In the knowledge graph, the entity is similar to the nodes in the graph, and the edge connecting each node is the relationship between the entities. Therefore, the most obvious advantage of using knowledge graph is that it can accurately describe data and their relationship, and manage knowledge quickly and accurately.

According to different research fields, knowledge graph can be divided into two categories: general knowledge graph and domain knowledge graph. Most of the large-scale knowledge graphs are general-purpose knowledge graphs, such as Knowledge Graph developed by google, ‘Zhixin’ knowledge graph constructed by Baidu, etc., but the research and application of knowledge graph in the industry field is less. Li Yang[4] constructed a knowledge graph oriented to medical information by analyzing medical information, so that medical data can be well managed and organized. And the medical knowledge graph is displayed visually. Pan Lihu[5] realized the construction of the knowledge graph of the coal mine field, effectively analyzed the knowledge of the coal mine field, and monitored the development of the coal mine field on the basis of the knowledge graph of the coal mine field. greatly enhance the safety of workers in the coal mine field. Wen Bilong[6] constructed the knowledge graph of reservoir geological domain in order to manage and share rich knowledge in the field of reservoir geology. Ren Haoli[7] have constructed a knowledge graph of the law of warship activity by studying and analyzing the law of warship activity, which can provide military intelligence information in the field of warship activity.

3 Knowledge system construction

This paper mainly carries on the ontology modelling for SOP domain, and on this basis, combined with the research on the construction method of the knowledge graph, completes the construction of the knowledge graph of SOP domain. The overall framework of the SOP

knowledge graph is presented in figure 1, by crawling the structured and semi-structured ship data, and then extracting and merging the data, so as to realize the construction of the knowledge graph.

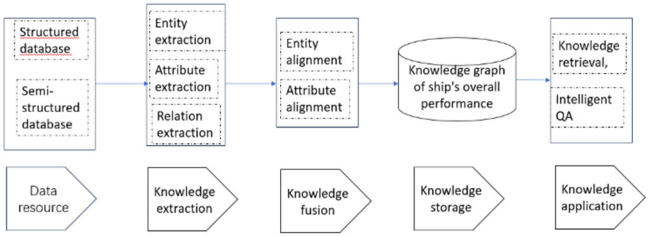

Fig. 1 General framework of knowledge graph construction of SOP domain

3.1 Data extraction

The primary task of constructing the knowledge graph of SOP domain is to collect data. According to its type, data can be divided into structured data, semi-structured data and unstructured data. Structured data refers to the data stored and represented according to a specific format, such as the data stored in MySQL and Oracle databases. Structured data can be directly read and used without other operations, and can be directly used to build a graph. Semi-structured data, including XML, HTML documents, excel tabular data and so on, need some simple data processing operations before they can be used for the construction of graphs. Unstructured packets contain a variety of incomplete data, such as documents, images, audio and video, which require complex data processing.

The data mainly comes from the data recorded by the China Classification Society Resource Center (https://www.ccs.org.cn/ccswz/),Ship data of the Cooperative Management platform of the Maritime Safety Administration of the people's Republic of China (https://ais.msa.gov.cn/index.html) ,VesselsFinder(https://www.vesselfinder.com/vessels) .

The ship data of the cooperative management platform of the Maritime Safety Administration of the people's Republic of China belongs to structured data, and the original data contains 4363 different ship information. The VesselsFinder platform contains 650977 ships, each of which contains data information such as ship name, ship type, flag, gross tonnage, year of completion and so on. There are a large number of semi-structured data in the ship record of China Classification Society. the original data contain the information of 3164 different domestic ships and 5947 different international ships. Each data includes the ship type, ship name, flag country, gross tonnage, depth, port of registry and other basic data of ship performance.

To crawl semi-structured data, we adopt some anti-crawler techniques, such as prohibiting cookie to prevent web stations from identifying users' information, compiling user agent pools to avoid exposing their identities, selecting different head; settings for each crawler to prevent frequent visits to websites in a short time, and so on. The crawled data is cleaned, format converted, and then the processed message is stored in the form of triple(<subject-predict-object>). A total of 318656 pieces of triple data were extracted.

3.2 Ontology construction

At present, there are a variety of ontology construction methods, including seven-step method, TOVE method and so on. Compared with other methods, seven-step method has a more complete life cycle and can achieve semi-automatic construction, and there are corresponding mature and detailed technical methods. This paper uses the seven-step method to build ontology, and uses protégé ontology editing tool to complete the construction of SOP domain ontology. The specific process is shown in figure 2.

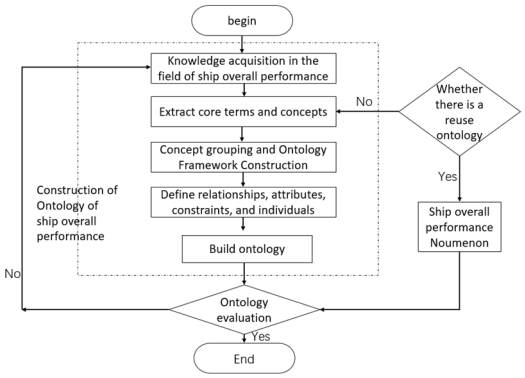

Fig. 2 Construction process of SOP ontology

(1) Analyze the fields and objects included in the SOP by flipping through the relevant books, relevant ship papers and other materials, abstractly summarize the entities, relations and attributes related to the SOP, in order to prepare for the construction of ontology.

(2) Search for whether there is a reusable ontology to investigate the existing open-source ontology models, understand the ontology entities, relationships and attributes and other information, and analyze whether it can be used in the SOP knowledge graph.

(3) Search and obtain knowledge in the SOP domain by surveying relevant experts, using the book 'Introduction to Ship Technology and Design'[8] as the main reference, relevant literature contained in CNKI, Wanfang and other ship related websites to gain knowledge.

(4) Extract the terms and concepts in the SOP domain to analyze and classify the data, and extract the outline of related classes.

(5) The concept is grouped and the ontology framework is displayed in the form of classes and subclasses, for example, ships are the first class, the second subclasses are military ships and civil ships, and the subclasses can inherit all the attributes of the ship parent class. In protégé, this paper defines six categories: App, manufacturing management, country, port, hull structure, ship and so on. Partial entities in SOP Knowledge graph are listed in table 1.

Table 1. Partial entities in SOP Knowledge graph

Entity name	Entity description	Secondary subclass
App	Ship-related App	
Manufacturing management	Ship manufacturing and management company	Shipyard, shipowner
Country	Countries in the five regions	Asia, Europe, America, Africa, Oceania
Port	The port at which the ship is located	Domestic ports and foreign ports
Hull structure	The components of a ship	Superstructure, main hull, deck components
Ship	Type of ship	Military ships, civil ships

(6) Defining relations, attributes, constraints and individuals to build the ontology in the SOP domain, we can not only define classes and hierarchical relations, but also need to define attributes, relations and so on. A relationship is used to connect two different entities, for example, the relationship 'belong-to' represents the flag state of the ship. Attribute is used to represent the data contained in an entity, for example, Haixun 1003, the shipyard of this ship is 193m. In protégé, relations and attribute distributions are described as object attributes and data attributes. Object attributes refer to some relationships between classes, and data attributes refer to the unique characteristics of a class. Partial object attributes and data attributes defined in the SOP knowledge graph are listed in table 2 and table 3.

Table 2. Object attributes defined in the SOP knowledge graph

Object attribute	Object attribute description	Domains	Ranges	Example
Belong-to	Country of registry	Ship	Country	< Fuxu, belong-to, French >
Construct	A ship's construction plant	Shipbuilding plant	Ship	< Guangzhou China Shipbuilding Huangpu Shipbuilding Co., Ltd, construct, Sea patrol 1003>
Ship-owner	The owner of a ship	Shipowner	Time	< Qinzhou Maritime Safety Bureau,Ship-owner, Sea patrol 1003>
Is-location	The port at which the ship is located	Ship	Port	< Pearl electricity dredging 1, Is-location, Zhuhai >
HasPart	The whole contains part	overall	part	< De Peng,HasPart, Upper deck >
PartOf	A part belongs to the whole.	part	overall	< Upper deck, Partof, De Peng >

Table 3. Data attributes defined in the SOP knowledge graph

Data attribute	Data attribute description	Mainly include	Domains	Ranges

Principal scale	Used to indicate the size and transport capacity of a ship.	Draught,vertical length, width, depth, length, freeboard, ship width, captain, design freeboard	ship	float
Time	Important time nodes related to ships	Launching time,ship completion time,shipfaketime,seaworthiness time	ship	date
Navigational performance	Have the ability to sail in waters	Buoyancy, stability, sinking resistance,rapidity,maneuverability, wave resistance, swing	ship	string
Ship tonnage	Used to indicate the size and transport capacity of a ship.	Empty shipweight,volume tonnage, weight tonnage	ship	float
Speed performance	The distance the ship sails relative to the sea floor in unit time.	Average speed, emergency speed, maximum speed, service speed, sea speed, port speed, etc.	ship	float
Ship form coefficient	Indicates the fatness and thinness of the underwater part of the ship	Volume coefficient, area coefficient	ship	float
Concealment	The performance of hiding the characteristic signal in the background		ship	string
Volume performance	The ability to indicate the volume of cargo the ship is carrying.	Cargo hold volume, deck area, special compartment	ship	string
App	Ship-related App	AppDescription,major,introduction, discipline, etc.	App	string
Ship-strength	The ability of the hull structure of a ship to resist various external forces without causing serious deformation or damage.	Total longitudinal, transverse, local and torsional strength	ship	string
Endurance	The maximum distance that can be reached at a given speed by burning at full load.		ship	string
Self-sustaining power	The number of days that fresh water and food are available on board.		ship	string

In order to complete an ontology model that can be used in practice, the last step is to add the corresponding instance to the ontology and complete the case filling. The data extracted from each ship data website is added to the SOP ontology.

4 Knowledge fusion

There are a variety of data sources for the SOP domain, and there may be many ways to record the same information, such as 'ship length'. Different websites have different descriptions, including 'length of ship', 'ship length' and so on. The extracted information is fused and cleaned to reduce the redundancy of the data as much as possible, which is more conducive to the construction of the SOP knowledge graph.

Entity and attribute alignment is to judge whether the entities or attributes described in different ways are the same or not, and unify the entities or attributes described in different ways into one. the process of aligning entities and attributes by calculating text similarity and using rule matching. In the SOP domain, there are many descriptions of various entities and attributes, such as country, ship name, ship performance and so on, and some of them include non-uniform writing methods such as English and abbreviations, which cause great difficulties in entity and attribute alignment. For example, the overall performance attribute of 'tonnage' is recorded in the ship catalogue query of China Classification Society, while in the book 'introduction to ship Technology and Design', this attribute is recorded as 'ship tonnage'. Alignment can be made between entities and attributes that point to the same entity through translation[9], text similarity calculation[10], and manual recognition. Examples of attribute alignment are listed in table 4.

Table 4. Examples of attribute alignment

Sample	Data source	Original attribute name	Align attribute name
1	Ship catalogue of China Classification Society	Gross ton	Ship tonnage
	《An introduction to warship Technology and Design》	tonnage	
2	Ship catalogue of China Classification Society	Draught (full load)	Full load draught
	《An introduction to warship Technology and Design》	Full load draught	

5 Mapping storage and application of knowledge graph

The knowledge graph stored in the Virtuoso graph database can be easily queried and modified, and the SOP knowledge graph can be visualized. There are four types of classes, relations, attributes and instances in SOP ontology, and there are three kinds of elements: nodes, relations and attributes in the graph database, so it is necessary to define specific mapping rules to map the ontology to the graph database for storage.

A. Nodes: the definition of nodes in the virtuoso graph database is similar to the entities and instances in the SOP ontology, such as ships, countries and so on. The classes and

instances are mapped to the nodes in the graph database, such as Sea Patrol 1003, China and other instance data is a separate node in Virtuoso.

B. Relationship: the different nodes in the virtuoso diagram database are connected by relationships to form a complete knowledge network. The SOP ontology contains the most basic inter-class relationship (subclass-of), class-instance relationship (instance-of), class-to-class relationship (belong-to) and so on. The relationships in these ontologies can be transformed into the relationships of nodes in the graph database.

C. Attributes: The SOP ontology contains object attributes and data attributes, in which object attributes are used to describe the relationship between classes, such as belong-to is to describe the relationship between the ship and the state, that is, the ship's registry country. Data attribute is to describe the internal attributes of a class, such as the average speed is to describe the average speed of the ship when sailing, which belongs to a data attribute of the ship itself. The object attributes in the ontology can be used as the relationship between nodes, and the data attributes can be saved in the Virtuoso graph database as node attributes.

6 Conclusion

This paper designs the ontology of the SOP domain by using the seven-step method, and then defines the mapping rules between the ontology and the Virtuoso graph database, and stores the ontology and instances data in the graph database, so that the SOP knowledge graph can be displayed constructed. The knowledge graph in the SOP domain can be applied to the actual scene to build an intelligent question answering system, so that ship researchers and engineers and technicians can quickly, conveniently and accurately search information and improve work efficiency.

References

[1] Deligiannis P. Ship performance indicator[J]. Marine Policy, 2017, 75: 204-209.

[2] Zou X. A survey on application of knowledge graph[C]//Journal of Physics: Conference Series. IOP Publishing, 2020, 1487(1): 012016.

[3] Fensel D, Şimşek U, Angele K, et al. Introduction: what is a knowledge graph?[J]. Knowledge graphs: Methodology, tools and selected use cases, 2020: 1-10.

[4] LI, Y., DU, R., ZHANG, H., (2022) Knowledge Graph Construction for Pharmaceutical Information. J. Computer Technology and Development., 32(10):189-193.

[5] PAN, L., ZHANG, J., ZHANG, Y., (2019) Construction of knowledge map in the coal mining field. J. Computer Application and Software., 36(08):47-54+59.

[6] WEN, B., XUE, G., (2021) Research on knowledge map construction in the field of reservoir geology. J. Computer Technology and Development., 31(12):204-210.

[7] REN, H., LUO, F., (2022) Research on knowledge graph analysis system for ship activity pattern. J. ship science and technology., 44(01):159-164.

[8] Shao Kaiwen, Ma Yunyi. Introduction to Ship Technology and Design, National Defense Industry Press,2013.10.

[9] Wang D, Wang S. SE-Former: Incorporating sentence embeddings into Transformer for low-resource NMT[J]. Electronics Letters, 2023, 59(11): e12840
[10] Wang J, Dong Y. Measurement of text similarity: a survey[J]. Information, 2020, 11(9): 421.