

Research on Representation and Retrieval for Equipment Fault Information Based on Domain Ontology

Yongjie Li^{1, a}, Houxiang Wang^{2, b} and Jingjing Zhang^{3, c}

^{1,2}Electronic Eng. College, Naval Univ. of Engineering, Wuhan Hubei 430033 China;

²Naval Engineering Design and Research Inst., BeiJing 100070 China

^acumtlyj@163.com, ^bwxh_hg@163.com, ^czjj_1977@163.com

Keywords: equipment maintenance; fault information; domain ontology

Abstract. Analyzed the shortcomings of fault information management in the field of warship equipment maintenance. An ontology-based method for equipment fault information modeling is proposed on the basis of the characteristics and the actual in warship equipment maintenance. The modeling process of equipment fault information ontology is introduced and the model is built using Protégé. In view of the lack of case retrieval based on keywords technology, the comprehensive retrieval algorithm based pre retrieval is designed and a prototype system is developed, validated and applied. The results showed that the algorithm can greatly improve the precision of fault case retrieval.

Introduction

A large number of maintenance knowledge are accumulated in the process of warship equipment maintenance and support. At present, the main using of the maintenance knowledge is building fault case systems in order to manage, retrieve and utilize fault information (called fault cases). However, there are lots of shortcomings in using fault cases such as the low effective management and the low ability of sharing, reusing and case reasoning because of the lack of unified canonical representation in different equipment fault information management systems.

For his better concept hierarchical structure, logical reasoning ability and ability to express knowledge, ontology technology has been extensively used in many knowledge management fields such as medical diagnosis, product design and manufacturing etc. [1]. To provide effective guidance on work of equipment maintenance and support, this paper applies ontology theory into the equipment fault case knowledge management, explores the modeling representation methods and normative structure for equipment fault case knowledge and researches on equipment fault case knowledge retrieval algorithm.

Fault case knowledge ontology construction

There are many ontology description languages such as SHOE, XOL, RDF, OIL, OWL. W3C launched RDF (S), DAML+OIL and OWL as the standard for ontology description language. OWL is used as the main ontology description language for its good compatibility, interactivity and rich semantic expression ability [2].

The process of establishing the equipment maintenance fault case knowledge ontology generally includes the following stages:

- 1) Determine the scope and target of equipment fault case knowledge ontology;
- 2) Collect and analyze equipment fault case knowledge;
- 3) Determine core concepts of equipment fault case knowledge;
- 4) Establish conception hierarchy of equipment fault case knowledge;
- 5) Define the concepts and attributes of equipment fault case knowledge;
- 6) Code and instantiate the equipment fault case ontology;
- 7) Examine and evaluate of equipment fault case ontology.

The works of equipment maintenance are not only related to the management and maintenance unit, maintenance personnel, spare part, detection equipment, maintenance standard, technical data and documentation, maintenance cost etc. but also associated with the ontology warship, equipment and equipment system. The equipment fault case ontology concepts can be defined as follows.

Definition 1: The fault case ontology of equipment maintenance, EMFCO for short. It can be defined as 9-tuple, $EMFCO \rightarrow \{I_C, D_C, T_C, E_C, M_C, P_C, C_C, H_C, R_C\}$.

Where I_C is basic information of fault case, D_C is the fault description, T_C is fault opportunity, E_C is fault influence, M_C is the fault mode, P_C is the fault parameters, C_C is the fault cause, H_C is the fault handling and R_C is the fault resources.

Definition 2: Basic information of fault cases ontology, I_C for short. It can be defined as 6-tuple. $I_C \rightarrow \{CID_I, SH_I, EP_I, PO_I, HT_I, RT_I\}$.

Where CID_I is the fault case ID, SH_I is the warship of fault; EP_I is equipment of fault; PO_I is the equipment system, HT_I is the fault happen time and RT_I is the fault repaired time.

Definition 3: Cause of fault case, C_C for short. C_C can be defined as 3-tuple. $C_C \rightarrow \{CG_C, MH_C, RA_C\}$.

Where CG_C is the fault category, MH_C is the failure mechanism and RA_C is the cause analysis.

Definition 4: Handling of fault case, H_C for short. It can be designed as 3-tuple. $H_C \rightarrow \{TM_S, QU_S, EV_S\}$.

Where TM_S is the troubleshooting methods, QU_S is maintenance quality and EV_S is the maintenance evaluation.

The concept and attribute modeling representation of equipment fault case used by Protégé-4.3 is shown as Fig. 1.

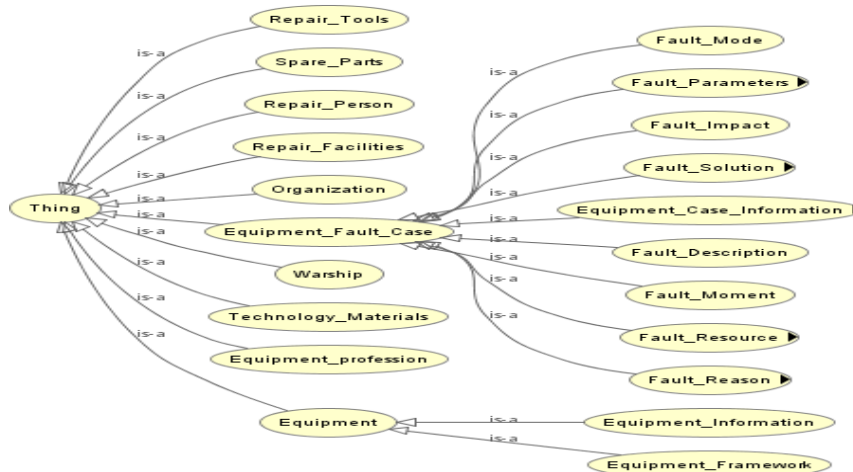


Fig. 1. Concept and attribute model of equipment fault case

In addition to the common Kind-of relation, Part-of relation and Equivalent-of relation, the relationships between the concepts of fault case knowledge ontology also includes other types of relationships, such as belong-to relationship between maintenance personnel and institution (inverse relation is Have-of), the Use-to relation between maintenance tools and equipment fault (inverse relation is Use-of), the Has-error relation between equipment and fault case (inverse relation Error-of), the Repair-of relation between maintenance personnel and equipment (inverse relation is Repaired-by) and the Consumption-of relation between spare parts and fault case (inverse relation is Consumed-by) etc..

Fault case knowledge retrieval

The retrieval and matching algorithms are very important for maintenance personnel can quickly retrieve the fault case knowledge. The equipment fault case knowledge involves conceptual information, numerical information and text description information. We can't accurately retrieve fault case knowledge only use one of them. So in this paper, we put forward a comprehensive fault case knowledge retrieval algorithm based on pre-retrieval.

In order to reduce the scale of fault case retrieval spatial and improve the efficiency of fault case retrieval, we conducted a preliminary filtering of fault case base through the method of matching the clear attributes of fault case, such as warship name, equipment professional, equipment name, equipmentsystem etc.. This paper uses SPARQL language to realize the fault case filtering[3].

Similarity calculation of fault case is divided into local similarity and the overall similarity. The local similarity is the similarity of a vector attribute between the fault case be retrieved and the fault case in case base. The overall similarity is the similarity value between the fault case be retrieved and the fault case in case base, which is calculated as the sum of weighted local similarity. The fault cases in case base are selected when it's overall similarity is greater than the threshold which is set by domain experts or system users[4].

The vector attribute in the fault case is mainly divided into numerical type, string type, enum, boolean type and ontology type. This paper focuses on the similarity calculation of digital type, string type and ontology type.

We suppose C_i is a fault case in fault case base which is indicated by n attributes, namely $C_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{in})$. Where x_{ij} is the j^{th} attribute of C_i . Q is the fault case be queried, where y_j is j^{th} attribute of Q corresponding to x_{ij} .

(1) Similarity calculation of numerical type feature value

We define y_j is a numerical attribute of a new fault case Q and x_{ij} is the corresponded attribute of a fault case in fault case base. There are three types of y_j value: definite number, interval number and fuzzy number.

1) When y_j is a definite number, the attribute similarity $Sim_i(s_j)$ can be calculated according to Eq. 1.

$$Sim_i(s_j) = \frac{1}{1 + d(C_{ij}, Q_j)} \quad (1)$$

$d(C_{ij}, Q_j)$ is the distance of two fault cases and its value is

$$d(C_{ij}, Q_j) = \begin{cases} \frac{|x_{ij} - y_j|}{\max(x_j) - \min(x_j)} & \max(x_j) - \min(x_j) \neq 0 \\ \frac{|x_{ij} - y_j|}{|x_{ij}|} & \max(x_j) - \min(x_j) = 0 \text{ and } x_{ij} \neq 0 \\ |x_{ij} - y_j| & \max(x_j) - \min(x_j) \neq 0 \text{ and } x_{ij} = 0 \end{cases} \quad (2)$$

$\max(x_j)$ and $\min(x_j)$ are the maximum and minimum values of x_j in fault case base.

2) When y_j is a interval number, the attribute similarity $Sim_i(s_j)$ can be calculated according to Eq. 3.

$$Sim_i(s_j) = \begin{cases} 1 & x_i \in y_j \\ \frac{1}{1 + d(C_{ij}, Q_j)} & x_i \notin y_j \end{cases} \quad (3)$$

$d(C_{ij}, Q_j) = \min(|x_{ij} - y_{ja}|, |x_{ij} - y_{jb}|)$, y_{ja} and y_{jb} are the endpoint values of y_j .

3) when y_j is a fuzzy value and it can be shaped like "relation operator + v", where v is the target value, the attribute similarity $Sim_i(s_j)$ can be calculated according to Eq. 4.

$$Sim_i(s_j) = \begin{cases} \frac{t - \min(f)}{v - \min(f)} & t \in [\min(f), v] \text{ and } t \neq v \\ 1 & t = v \\ \frac{\max(f) - t}{\max(f) - v} & t \in [v, \max(f)] \text{ and } t \neq v \\ 0 & \text{others} \end{cases} \quad (4)$$

t is the true value of x_j in fault case base.

(2) Similarity calculation of string type feature value

In order to improve the matching accuracy, string feature vector can be constructed and its elements are the frequency of a certain keywords in the string. The similarity can be calculated by calculating the two vector distance.

We define x_{ij} and y_j are the values of the two strings be compared. $K = \{k_1, k_2, k_3, \dots, k_m\}$, the keyword k_i is composed of ontology vocabularies and terminologies.

$V_1 = \{xv_1, xv_2, xv_3, \dots, xv_m\}$, V_1 is the feature vector constructed by x_{ij} , xv_i is the frequency of k_i in x_{ij} .

$V_2 = \{yv_1, yv_2, yv_3, \dots, yv_m\}$, V_2 is the feature vector constructed by y_j , yv_i is the frequency of k_i in y_j .

The similarity $Sim_i(s_j)$ can be calculated according to Eq. 5.

$$Sim_i(s_j) = \sum_{i=1}^m xv_i \times yv_i / \sqrt{\sum_{i=1}^m xv_i^2} \sqrt{\sum_{i=1}^m yv_i^2} \quad (5)$$

(3) Similarity calculation of ontology type

The fault case knowledge are classified and defined as tree structure in the previous chapters. Each concept and instance is a tree node of the fault case knowledge ontology. When the attribute value is the ontology instances, the similarity can be calculated by calculating the distance of the node in tree structure, as shown in Eq. 6.

$$Sim_i(s_j) = \begin{cases} \frac{2 \times Dep(Msc(x_{ij}, y_j))}{Dis(x_{ij}, y_j) + 2 \times Dep(Msc(x_{ij}, y_j))} & x_{ij} \neq y_j \\ 1 & x_{ij} = y_j \end{cases} \quad (6)$$

$Msc(x_{ij}, y_j)$ is the nearest father node of x_{ij} and y_j . $Dep(Msc(x_{ij}, y_j))$ is node depth of $Msc(x_{ij}, y_j)$ and $Dis(x_{ij}, y_j)$ is the shortest distance between x_{ij} and y_j (Minimum number of nodes from x_{ij} to y_j).

(4) The overall similarity calculation of fault case

The overall similarity $SSim(C_i, Q)$ between Q and C_i can be calculated according to Eq. 7.

$$SSim(C_i, Q) = \frac{\sum_{j=1}^n (sw_j \times Sim_i(s_j))}{\sum_{j=1}^n sw_j} \quad (7)$$

sw_j is the weight of j^{th} attribute which can be determined by field experts or the AHP (Analytic Hierarchy Process) algorithm [5].

Application example

The fault case knowledge ontology is modeled by protégé 4.3. The attributes and relations between them are designed. More than 150 instances of fault case knowledge are extracted through the collection and arrangement of equipment maintenance materials of a naval ship troops in five years.

We build an prototype system using Eclipse, Java JDK and Jena as development platform and environment, Oracle 11g as background database. Based on the communication with domain experts, we set the weight of fault influence, fault mode, fault parameters, fault causation, fault solution etc. and fault case similarity threshold is 0.65. Compared with the traditional keyword retrieval model, the recall and precision of fault case are greatly improved.

Summary

The concepts, attributes and the relations between them in equipment fault case knowledge are studied based on the analysis on the characteristics in the field of warship equipment maintenance and support. The knowledge model is built using Protégé. In view of the lack of case retrieval based on keywords technology, the comprehensive retrieval algorithm based pre-retrieval was designed and a prototype system is developed, validated and applied. The results showed that the algorithm can greatly improve the comprehensiveness and precision of fault case retrieval. The study in this paper can provide technical method and basis for the equipment maintenance knowledge sharing, and improve the knowledge's sharing and reuse ability.

References:

- [1] Yu Xu, Liu Ji-hong, He Miao. Computer Integrated Manufacturing Systems. Vol.02(2011), p.225-231. (In Chinese)
- [2] Titus Schleyer; Melissa Castine and Miguel Humberto Torres-Urquidy: J Am Dent Assoc. Vol.142(2011), p.250-251.
- [3] Carlos Buil-Aranda; Marcelo Arenas; Oscar Corcho; Axel Polleres: Web Semantics: Science, Services and Agents on the World Wide Web. Vol.18(2014), p.1-17.
- [4] Delia Rusu; Blaž Fortuna; Dunja Mladenić: Applied Ontology. Vol.9(2014), p.65-95.
- [5] Jin Juliang, Wei Yiming, Pan Jinfeng: System Engineering-theory & Practice, Vol.01(2004), p. 63-69. (In Chinese)