

Knowledge extract and ontology construction method of assembly process text

Yimin Du¹, Lingling Shi^{1,*}, Xiang Zhai², Hanqing Gong¹, and Zhijing Zhang¹

¹School of Mechanical Engineering, Beijing Institute of Technology, Beijing, China

²Beijing Institute of Electronic System Engineering, Beijing, China

Abstract. The actual product assembly process mainly relies on manual assembly by workers, and the personal experience of workers is difficult to effectively reuse. Ontology as a knowledge management and expression tool is gradually applied in the field of assembly. However, the manual construction of the ontology is time-consuming and labor-intensive, and the automatic construction of the ontology requires a large number of corpora for training, both of which are difficult to obtain a good assembly case ontology. This paper proposes a method in which automatically extracts relevant knowledge from case assembly process files to generates case database and integrates ontology framework of assembly domain to construct ontology. It shows that the accuracy can be guaranteed on the basis of the rapid construction of case ontology. The feasibility of this method is proved by a practical case.

Keywords: Ontology construction, Knowledge extract, Assembly case ontology.

1 Introduction

At present, case-based assembly process generation mainly relies on skilled workers. This traditional process design method relies too much on workers and consumes a lot of labor. ^[1] There is no scientific management and effective use of existing knowledge. In order to better manage the knowledge accumulated in the assembly process and realize the rapid generation of the case assembly process, ontology technology as a knowledge representation, management, and reuse technology began to be gradually used Applied in the field of assembly.

Ontology was first proposed in the field of philosophy as a philosophical concept describing the essence of things. ^[2] With the continuous development, ontology has been endowed with other meanings. The definition proposed by Studer. ^[3] that ontology is a clear, formal and normative description of the shared conceptual model is now widely accepted. It describes the meaning of the four sides: 1. Shared, it means that the knowledge in the ontology is universally recognized. 2. Conceptualization, which means that the ontology

* Corresponding author: lshi@bit.edu.com

represents the description of things as a set of concepts. 3. Clear, it means that all the conceptual attributes and axioms in the ontology are clearly defined. 4. Formal, means that the ontology language is a readable and processable form of a computation.^[4-6] Ontologies can be divided into general ontology, domain ontology, application ontology, etc. according to the different fields in which they are applied.^[7]

There are three main ways of ontology construction, manual, automatic and semi-automatic construction.^[8-9] Manual construction mainly clarifies the concept and relationship of ontology through domain experts, and the commonly used methods include the seven-step method^[10], IDEF5^[11], and methontology.^[12] The manually constructed ontology is more professional and highly reliable, but it takes a lot of time.^[13] The automatic construction of ontology has gradually become a research hotspot due to its effectiveness. Zuo Hongtao,^[14] Liang shuqin^[7] used BP neural network to extract the concept of ontology, and obtained the relationship among the concepts through K-means clustering to realize the automatic construction of water environment ontology. Fatima N^[15] calculates fuzzy number based on concept similarity and uses it to extract concepts and classification relations to construct ontology. The automatic construction of ontology based on deep learning often requires large-scale training corpus, while the sample size in the assembly field is too small to obtain a good training model.

In order to quickly generate assembly case ontology, this paper proposes an ontology construction method that automatically extracts case entity concepts and related attributes from case assembly process text information and merges assembly domain ontology template framework to generate assembly case ontology. For the assembly process files of different cases, the process content, work step content, assembly requirements, part information etc. are extracted to form the assembly process text, and the named entity recognition technology based on conditional random field [16-18] is used to realize automatic extraction of case concepts. The extracted entity concept and related attributes of assembly case are stored in an Excel template to form an assembly case database. Through the mapping rule [19], the case database is mapped to the assembly domain ontology conceptual framework to generate the assembly case ontology, which is visualized in the software Protégé [20]. The framework is shown in Figure 1.

The structure of this paper is as follows. The first part mainly introduces the concepts related to ontology and the research on ontology construction methods. The second part introduces the construction process of assembly domain ontology framework. The way to extract the concept of assemble case and related attributes are introduced in the third part. The generation and visualization of assembly case ontology are explained in the fourth part. In the fifth part, the feasibility of the construction method is proved by analyzing the concrete assembly cases.

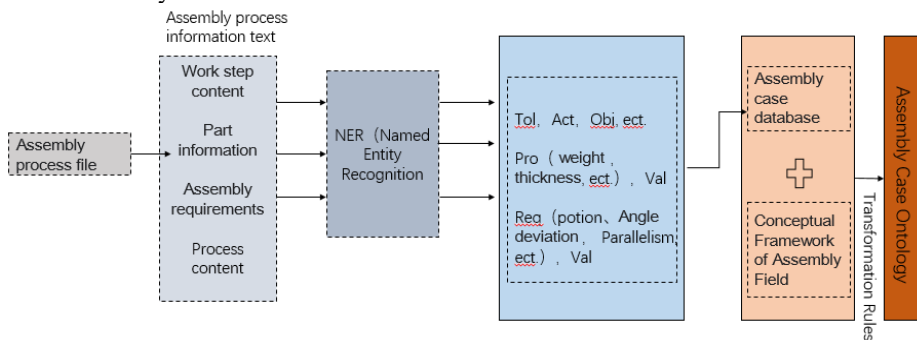


Fig. 1. Assembly case ontology construction process.

2 Construction of ontology conceptual framework in assembly field

According to the application classification of ontology in different fields mentioned above, concepts in the case assembly ontology are correspondently divided into top-level concepts, domain concepts and application concepts from top to bottom [21] in this paper. The conceptual framework of assembly domain ontology mainly includes the top-level concept and the domain concept. Application concept information is not included in this framework because the specific parts contained in different assembly cases are quite different, and the actual attributes and assembly requirements of different parts are also different.

The top-level concept mainly includes four categories: assembly abstract class, assembly concrete class, assembly time class, and assembly mark class. The assembly abstract class describes the concepts abstracted in the assembly process, including concepts such as assembly cases, assembly requirements, assembly dimensions, etc. The assembly concrete class describes the concept of objects objectively existing in the physical world during the assembly process, including assembly tools, assembly objects, assembly object characteristics, etc. Assembly time class describes concepts with time or sequence and sequence meaning, including assembly action, assembly process, assembly content, etc. The assembly marking class describes the content that needs to be specially marked in the assembly situation, including assembly time, assembly environment, etc.

The domain concept is the refinement and expansion of the top-level concept. For example, under the assembly concrete class, assembly tools can be divided into manual assembly tools and automatic assembly tools. As for the application concepts and instantiation information in the assembly case ontology will be implemented in the third part. Combining the above-mentioned assembly domain top-level concept and domain concept construction method, construct the ontology in Protégé, and realize the visual expression of the domain ontology conceptual framework. As shown in Figure 2.

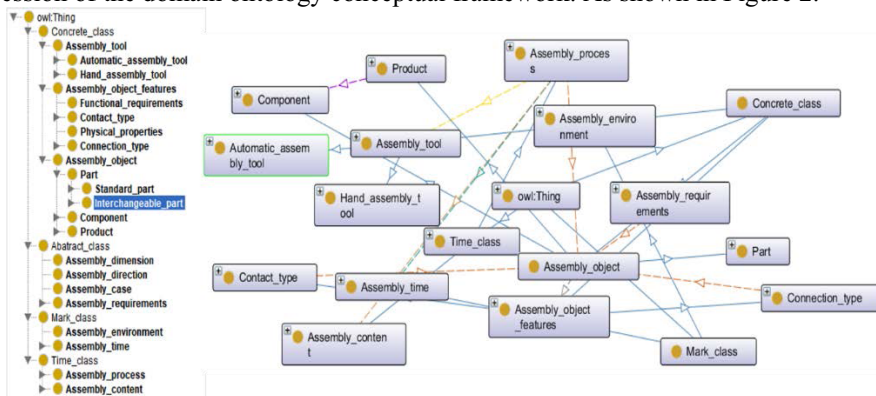


Fig. 2. Conceptual framework of assembly domain ontology.

3 Construction of ontology conceptual framework in assembly field

The extraction process for case concepts and partial concept attributes is shown in Figure 3. (1). Preprocessing; The text preprocessing is to ensure the accuracy of word segmentation results. In this paper, a thesaurus oriented to the assembly field is constructed for word segmentation, mainly in the following two ways: a) Use the crawler to get the domain approximate words; Set the initial assembly word set, use crawler technology to crawl the

synonym set of each word in the initial assembly word set, and add it to the assembly word set, and then crawl its synonym set from the newly obtained synonym set. To ensure semantic relevance, the number of iterations is set to 2. b) According to the actual case assembly process text, set new word extraction rules suitable for this task, and add new words to the assembly word database. For example, in the assembly process of cavity target, there are multiple marking piece parts. In order to distinguish their differences, the parts will be named according to the naming rule of "class name + number", such as "marking piece 01", "marking piece 02", etc.

Based on the construction of thesaurus, NLP-IR word segmentation is used to segment the assembly text, and the stop word list of HIT is used to filter the assembly text. (2) Annotation; The annotation method of BIEO is adopted to annotate the data into a format that can be processed by the computer, in which "B" represents the beginning of the entity, "I" represents the middle part of the entity, "E" represents the end of the entity, and "O" represents the non-entity part of the text. (3) Model training and testing; For the labeled data, divide the training set and the test set at a ratio of 7:3, input the training set to train the Conditional Random Field (CRF) model, the test set for testing, and the model outputs the labeled sequence. (4) Recombination; According to the output annotation sequence, the named entity is extracted by recombination. Taking the specific assembly sentence statement as an example, the statement processing result is shown in Figure 4

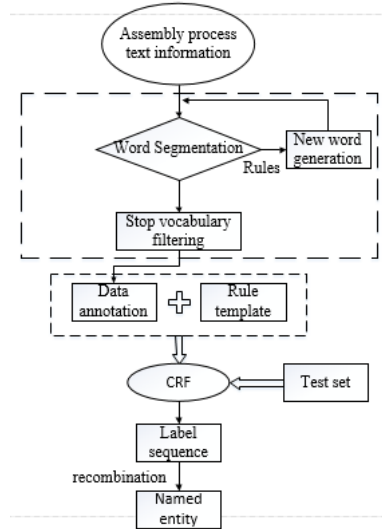


Fig. 3. The extraction process of assembly case concept and some concept attributes.

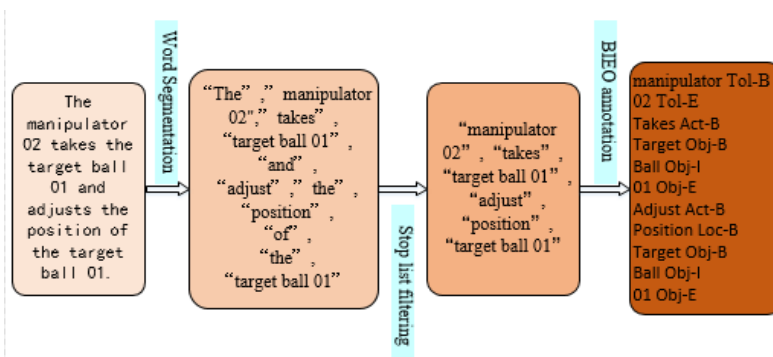


Fig. 4. Text processing procedure.

4 Generation of assembly case ontology

The Cellfie module is built in Protégé. By writing conversion rules in Cellfie, the mapping of Excel template data to the assembly case ontology framework can be realized to construct the assembly case ontology. The transformation rules are written by MappingMaster, which is a domain-specific language (DSL) based on Manchester OWL Syntax.

Take the relationship between the parts contained in the cavity target as an example. By defining the following MappingMasterDSL statement, the Cavity target can be established as a subclass of Product based on the data in the table, and has an individual Cavity target01. Cavity target01 has the physical attribute is_component_of, and is connected to the 6 individuals in the third column of the table through the is_component_of attribute. In this way, the mapping and conversion of the Excel template data to the ontology is constructed. As shown in Figure 5. In the MappingMasterDSL statement, "Class:" and "Individual:" are statements for generating classes and individuals, respectively. "Facts:" are statements for adding attributes. "@B2" refers to the content in cell B2. "@C*" refers to the content in column C, where the start and end rows can be restricted in the Cellfie template.

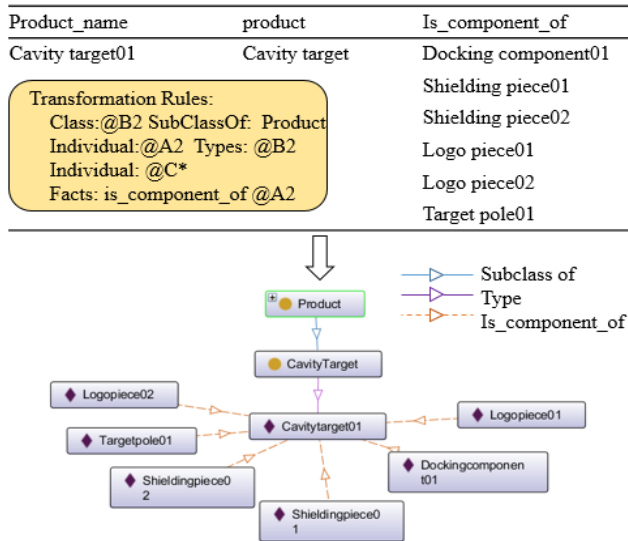


Fig. 5. Mapping of Excel template data to ontology.

5 Results and case analysis

5.1 Results

Due to the limitation of professional terms in the assembly field, the existing training corpus is very different from the assembly text. This article constructs a CRF training corpus from the assembly process description texts of three actual assembly cases, which includes a total of 288 short sentences and about 1800 basic tokens. The entity types are divided into assembly tool (Tol), assembly action (Act), assembly object (Obj), assembly location (Loc), assembly base (Bas). Using "BIEO" labeling method, and there are 16 kinds of labels in total. Use precision, recall rate, and F1 value as model evaluation indicators.

The precision, recall rate and F1 value of each label are shown in Table 1.

Table 1. The precision, recall rate and F1 value of each label.

Entity type	P	R	F1
Tol-B	1	0.97297	0.98630
Tol-I	1	0.97727	0.98850
Tol-E	0.97297	0.97297	0.97297
Act-B	1	0.95745	0.97826
Act-I	1	1	1.00000
Act-E	1	0.948718	0.97368
Obj-B	0.94739	0.81818	0.87806
Obj-I	0.82143	0.71875	0.76667
Obj-E	0.89474	0.77273	0.82927
Loc-B	0.78947	0.9375	0.85714
Loc-I	0.73438	1	0.84685
Loc-E	0.84211	1	0.91429
Bas-B	1	0.5	0.66667
Bas-I	1	0.5	0.66667
Bas-E	1	0.5	0.66667
O	1	0.96552	0.98246

5.2 Case analysis

For the cavity target, the assembly step description text was obtained from the assembly process document of the cavity target, and the application concept, assembly object, assembly action and other instance information as well as assembly step process number information were extracted from the assembly text of the cavity target by using the trained CRF model. The information obtained above is automatically stored in the corresponding position of the established Excel template to form the assembly case database, and the transformation rule library is established through the MappingMaster(DSL) statement to construct the mapping of the assembly case database to the domain ontology framework. As shown in Figure 6, the specific classes of cavity target, such as half-cavity class, gripping membrane class, target ball class, etc. are added to the existing domain ontology framework through mapping, namely the application concept of assembly case ontology. These classes are supplemented with instance numbers and operation procedure numbers.

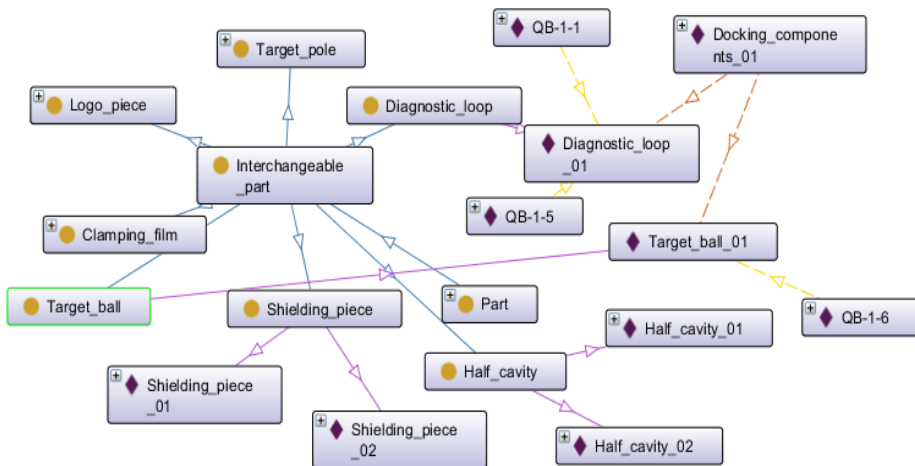


Fig. 6. Assembly case ontology.

6 Results and case analysis

This paper proposes an ontology construction method that automatically extracts case entity concepts and related attributes from case assembly process text and merges assembly domain ontology template framework to generate assembly case ontology. It is proved to be feasible with the case of cavity target assembly. Compared with the traditional manual ontology, this method automatically extracts application concepts and individual from the assembly process text through CRF, forms the assembly case database from the extracted data Excel template, and enriches the ontology framework through the mapping method, thereby improving with rapidity and versatility. The manual construction of the assembly domain ontology framework can ensure the accuracy of the assembly case ontology to a certain extent. However, there still are some limitations. For an actual assembly case, the information contained in the assembly process documents is limited, and the training corpus in the assembly field is so less that the model may have overfitting phenomenon. Therefore, in the following research, we will further integrate other level information of the assembly case to further supplement the assembly case ontology. As for the problem of named entity recognition with few samples, we consider to further optimize the model training results through data enhancement of the samples.

References

1. Liu, Jianhua, Qingchao Sun, Hui Cheng, Xiaokang Liu, Xiaoyu Ding, Shaoli Liu, and Hui Xiong. 2018. "The State-of-the-Art, Connotation and Developing Trends of the Products Assembly Technology." *Journal of Mechanical Engineering* 54(11):2–28.
2. Maedche, Alexander, and Steffen Staab. 2001. "Ontology Learning for the Semantic Web." *IEEE Intelligent Systems* 16(2):72–79.
3. Studer, Rudi, V. Richard Benjamins, and Dieter Fensel. 1998. "Knowledge Engineering: Principles and Methods." *Data & Knowledge Engineering* 25(1–2):161–97.
4. Gruber, T. R. 1993. "A Translation Approach to Portable Ontology Specifications." *Knowledge Acquisition* 5(2):199–220.
5. O'Leary, Daniel. 2005. "Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce."
6. Neches, Robert, Richard E. Fikes, Tim Finin, Thomas Gruber, Ramesh Patil, Ted Senator, and William R. Swartout. 1991. "Enabling Technology for Knowledge Sharing." *AI Magazine* 12(3):36.
7. Liang, Shuqin. 2016. "The Research and Application of Water Environmental Ontology Construction and Organization Methods." *Huazhong University of Science and Technology*.
8. Dong, Jin, Jian Wang, and Zhaoping Wang. 2021. "Automatic Ontology Construction for Human-Cyber-Physical Data Fusion in Manufacturing Domain." *Control and Decision* 1–7.
9. Ren, Feiliang, Jikun Shen, Binbin Sun, and Jingbo Zhu. 2019. "A Review for Domain Ontology Construction for Text." *Chinese Journal of Computers* 42(3):654–76.
10. Du, Wenhua. 2005. "Comparative Study on Ontology Construction Methods." *Journal of Information* 24(10):24–25.
11. Peraketh, Benjamin, Christopher P. Menzel, Richard J. Mayer, Florence Fillion, and Michael T. Futrell. 1994. *Ontology Capture Method (IDEF5)*.

12. Fernández-López, Mariano, Asunción Gómez-Pérez, and Natalia Juristo. 1997. "Methontology: From Ontological Art towards Ontological Engineering."
13. Ren, Ni, Yiwei Sun, Tong Bao, and Ting Guo. 2021. "Research on Ontology Construction Method in Agriculture: Case Study of Tomato Diseases and Insect Pests." *Information Research* 1(4):1.
14. Zuo, Hongtao. 2016. "Research about Concept Extraction and Mapping of Water Environment Ontology Based on Artificial Neural Network Algorithm." Huazhong University of Science and Technology.
15. Al-Aswadi, Fatima N., Huah Yong Chan, and Keng Hoon Gan. 2020. "Automatic Ontology Construction from Text: A Review from Shallow to Deep Learning Trend." *Artificial Intelligence Review* 53(6):3901–28.
16. Jiang, Huixing. 2012. "Research on Chinese Named Entity Recognition." Beijing University of Posts and Telecommunications.
17. Hu, Hongping, and Hui Zhang. 2008. "Chinese Named Entity Recognition with CRFs: Two Levels." Pp. 1–6 in 2008 International Conference on Computational Intelligence and Security. Vol. 2.
18. Kim, Seokhwan, Yu Song, Kyungduk Kim, Jeong-Won Cha, and Gary Geunbae Lee. 2006. "Mmr-Based Active Machine Learning for Bio Named Entity Recognition." Pp. 69–72 in Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers.
19. Hazber, Mohamed A. G., Ruixuan Li, Xiwu Gu, Guandong Xu, and Yuhua Li. 2015. "Semantic SPARQL Query in a Relational Database Based on Ontology Construction." Pp. 25–32 in 2015 11th International Conference on Semantics, Knowledge and Grids (Skg).
20. Noy, Natalya Fridman, Ray W. Ferguson, and Mark A. Musen. 2000. "The Knowledge Model of Protege-2000: Combining Interoperability and Flexibility." Pp. 17–32 in International Conference on Knowledge Engineering and Knowledge Management.
21. Wang, Hongyu, Chaohong Yang, and Yuwei Zhou. 2020. "Research on Content and Method of Military Domain Ontology Construction." *Information & Communications* 10:149–51.