

# A method for extracting design rationale knowledge based on Text Mining

Jihong Liu<sup>1,\*</sup>, Jiaji Wang<sup>1</sup>, and Kejian Wang<sup>1</sup>

<sup>1</sup> School of Mechanical Engineering and Automation, Beihang University, 100191 Beijing, China

**Abstract.** Capture design rationale (DR) knowledge and presenting it to designers by good form, which have great significance for design reuse and design innovation. Since the 1970s design rationality began to develop, many teams have developed their own design rational system. However, the DR acquisition system is not intelligent enough, and it still requires designers to do a lot of operations. In addition, the existing design documents contain a large number of DR knowledge, but it has not been well excavated. Therefore, a method and system are needed to better extract DR knowledge in design documents. We have proposed a DRKH (design rationale knowledge hierarchy) model for DR representation. The DRKH model has three layers, respectively as design intent layer, design decision layer and design basis layer. In this paper, we use text mining method to extract DR from design documents and construct DR model. Finally, the welding robot design specification is taken as an example to demonstrate the system interface.

## 1 INTRODUCTION

DR knowledge is an explanation of why a product is so designed [1, 2]. DR knowledge includes explicit knowledge (such as design specifications, operating instructions, patent documents, etc.) and invisible knowledge. Explicit knowledge generated by the designer during the design of the workpiece. Invisible knowledge generated in the minds of the designer. Since the 1960s, DR began to develop, it has been recognized as providing important information and knowledge for design reuse, design reasoning and design evaluation [3]. For example, DRed (Design Rationale editor) developed by Bracewell (University of Cambridge Engineering Design and Research Center), by recording graphic design rational elements and content to establish a design rational model and support designers to browse the design information and access. DR is widely used in machinery, construction, software development and other fields.

In order to better manage and use the DR knowledge, many studies are devoted to the establishment of DR knowledge base. There are two sources of DR, one is to capture DR in the design process, and the other is to capture the DR from the design files that have been formed. Nearly 20 years, the DR acquisition method research mainly focuses on capturing the DR that generated during the design process. This method can record the explicit knowledge and tacit knowledge produced by engineering designers in detail, but need to spend a lot of time and energy. At present, there is relatively little research on DR captured in design documents. DR documents( such as design specifications, test report, patent document) because of its complex

structure, complicated language, natural language is not easy to be computer recognition and other reasons have not been as an important source of DR. However, the design specifications, test reports, patent documents contain a lot of DR knowledge, for design reuse, design reasoning, design evaluation is very important [4]. Therefore, if there are some methods can accurately and efficiently extract DR from the design document, for the DR knowledge base, as well as the application of DR will have a great role in promoting. In view of the above mentioned problems, this paper proposes a method to extract the DR from the design document according to the Intent-Driven DR Knowledge model (IDDR) proposed by the team in the previous study. This method can improve design knowledge reuse efficiency.

Part one of this article describes the previous relevant research about DR. The second part puts forward the hierarchical model of design decision context. Part three presents an algorithm for obtaining DR from design documents and uses hierarchical model of design decision context to express. The fourth part uses the welding robot design manual as an example to demonstrate the system. The last part is the article summary.

## 2 DESIGN RATIONALE KNOWLEDGE HIERARCHY MODEL

A perfect DR model has a positive effect on the reuse of DR knowledge. At present, according to the different expression of DR, DR model is divided into two categories. The first approach is argumentation-based representation. In the argumentation-based DR

\* Corresponding author: ryukeiko@buaa.edu.cn

representation model, the design process is taken as a problem solving process and it is composed of multiple subprocesses [5]. The argumentation-based representation approach is represented by IBIS (issue-based information system) which is the most mature expression in the field of DR. The IBIS model originated from Kunz and Rittel, and was later extended to different models such as PHI (Procedural Hierarchy of Issue), QOC (Question-Option-Criteria), DRL (Decision Representation Language) [6-8]. However, the argumentation-based DR model still has the shortcomings: first, the model divorce the product description, its expression is only for the design decision-making process, can not express the design itself; second, the successful modeling process is based on the designer who have an accurate and full understanding of design issues, therefore it does not apply to all levels of the designer. But in the actual design process, the designer's understanding of the design problem is a process from scratch, the design problem in most cases is also a pathological problem, with an unclear initial state, unclear purpose and uncertain strategy. Therefore, the argumentation-based DR model can not fully reflect the typical characteristics of design activities such as design cognition and design iteration.

Intention-Driven Design Rationale knowledge modeling studies suggest that the designer's intentions interact with the environment, and co-evolution with it [9]. Intention-driven DR knowledge modeling can preserve design history by recording design intent evolution trajectories. Ganesha takes the design intent as a result of the continuous evaluation and refinement of the design process as a design rationale knowledge [10]. However, the current intention-Driven design rationale knowledge modeling research only concerned about the change process of the design intent, did not consider the design basis, reason and other design elements of design knowledge. Therefore the knowledge structure is simple, can not obtain comprehensive design rationale knowledge.

Our team based on the intention-driven ideas, put forward the design rationale knowledge hierarchy (DRKH) model. The DRKH model consists of three levels: design intention, design decision and design basis, as shown in Fig.1.

Design intent is the goal, plan and purpose of the designer to design thinking. It can explain the factors that affect the designer's problem solving, decision making and operation execution in the design process. Design intent is both a motivation for triggering designer thinking, and sometimes a direction for designers to think about development. Design intent can be decomposed into multiple levels of sub intent, when the realization of all sub intents, the parent's intent can achieve [11]. There is no need to re-decompose and can implement directly, called meta intent. Design decision refers to the designer referring to the relevant basis, according to the evaluation criteria, put forward, analyze and compare a number of design options, and get the final solution to achieve the design intent. The design basis is used to explain the reasons for decision making,

including the basis, standards, and tradeoffs of the designer in terms of his own expertise, experience, preferences, and situational information.

DRKH model compared with other models, have no perplexing relationship, but can clearly and completely express a product design process from scratch. It can reflect cognitive behavior such as thinking, imagination, and decision making in the designer's mind throughout the design process. Therefore, the method proposed in this paper is based on DRKH model.

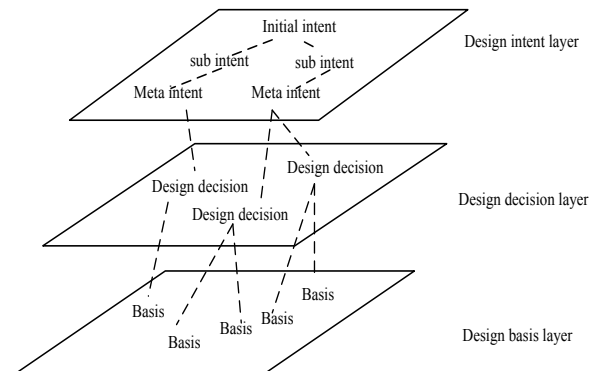


Fig. 1. Design rationale knowledge hierarchy model

### 3 DESIGN RATIONAL KNOWLEDGE EXTRACTION METHOD

#### 3.1 Design intent layer generation

The task of design intent generation is to extract phrases that can express the designer's design motivation. After a lot of research and analysis, in the design document, design intent generally exist at different levels of the title, and the title of the higher level, the corresponding design intent in the design intention level at the higher level, so the extracted design document's title can be used for building design intent tree. The design document exists in the form of WORD, PDF format, so use the JAVA POI Library in Apache to identify and extract the title of the design document. The algorithm flow is shown in Fig.2.

Step 1: Design document preprocessing.

- 1) Input the design document W;
- 2) See if the design document W has a standard title format. If there is no standard title format, precede step 3, and if so, precede step 4;
- 3) Use the Apache POI to change the design document W to make its title standardized;
- 4) Use the Apache POI method to get all of the titles in the document and the inclusion relationships that exist between them;
- 5) All the titles obtained will be stored in the design intention candidate P according to their level;

Step 2: Design intent extraction.

The title stored in the design intent candidate set P is not all intended as a design intent, so it is necessary to delete the title that is not design intent.

1) Construct a word set D, in which words can not be used as design intents after a large number of experiments, and similar words are stored in dictionaries and stored in D;

2) Matches the title in the design intent candidate set P with the vocabulary in the word set D. If the title contains the words in the word set D, it is deleted and the design intent set Q is obtained;

Step3: Design intent layer generation.

After step 1, step 2, the design intent set is preserved in the Q, the title and its relationship between the header and their relations that can be used as the design intent layer are generated. Therefore the design intent layer is generated according to the Q.

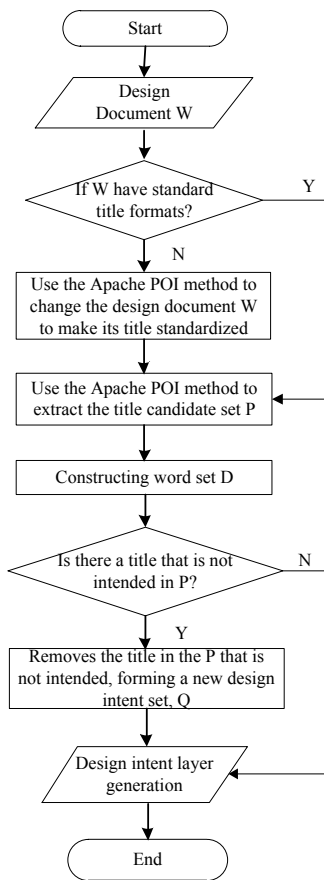


Fig. 2. Algorithm for design intent layer generation

### 3.2 Design decision layer generation

The design decision generation task is to extract design decision that complete a design intent. According to a large number research, features of the design decision statement is a verb plus noun form, namely verb object phrase, so that the design decision extraction is to extract relevant verb phrases related to design intent. In this paper, phrase structure based syntactic analysis is applied to extract verb object phrases from design decisions. The algorithm flow is shown in Fig.3.

Step 1: Design document preprocessing.

1) Input the design document W;

2) Divide design document W into n sentences  $S = (x_1, x_2, \dots, x_n)$ ;

3) Segment each sentence in S into words and conduct part of speech tagging;

4) Using Apache POI method to determine the range of design decisions that will extraction, the design intention layer extraction algorithm mentioned each design intention corresponds to a title, therefore, the content between adjacent titles is the design decision and basis for realizing the design intention;

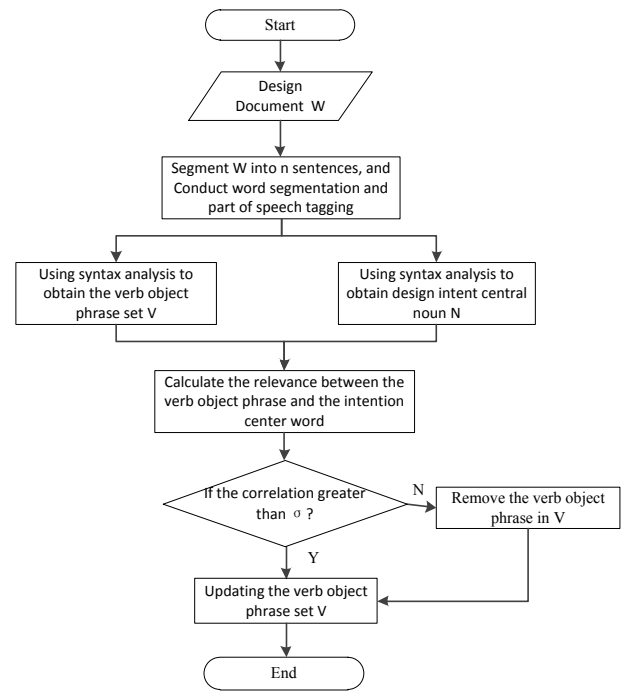


Fig. 3. Algorithm for design decision layer generation

Step 2: Design decision extraction;

1) Phrase structure based syntactic analysis is used to obtain the central word N of the design intent and all verb object phrases which are in the paragraph of a design intent. Put the verb object phrase into V, which is the candidate set of design decision;

2) Taking into account the design decision candidate set V obtained in 1), not all verb object phrases can be used as design decisions, so the verb object phrase that is not a design decision should be deleted from the V. In this paper, we adopt the method of correlation calculation, calculates the correlation between the words contained in adjacent titles and the relevance of the design intent central word N. Set the threshold  $\sigma$ , extract the relative degree is greater than the threshold  $\sigma$  word set C, match the verb object phrase V with the word set C. The verb object phrase that contains any of the words in the word set C is the design decision corresponding to the design intent. Calculation formula of correlation (1):

$$s(i, j) = \frac{tf(i, j)}{w(i, j)} \quad (1)$$

s(i, j) represents the degree of correlation between words i and j; tf(i, j) represents the number of times that the word i, j appears in a sentence at the same time; w(i, j) represents the relative distance of i, j, see equation (2):

$$w(i, j)$$

$$= \frac{\text{The number of characters between } i \text{ and } j}{\text{The number of characters contained in the content}} \quad (2)$$

3) Delete the verb moving object phrases that are not design decisions in the design decision candidate set.

Step 3: design decision layer generation.

1) Update design decision candidate set V, that is the design decision set;

2) The verb object phrase of the design decision is connected with the corresponding design intention to form a design intention-design decision hierarchy.

### 3.3 Design basis layer generation

The design basis is used to explain the reasons for the decision, including the basis, the standards, and the tradeoffs that the designer makes on his own expertise, experience, preferences, and situational information. The design basis is explanatory sentences, so the design basis sentence usually contain "because....." "as a result of....." "on the basis of....." and other explanatory words and phrases In this process, we draw on the related algorithms of text mining. The algorithm flow is shown in Fig.4.

Step 1: Design document preprocessing.

1) Input the design document W;

2) Divide design document W into n sentences

$$S = (x_1, x_2, \dots, x_n);$$

3) Establish a design basis vocabulary library O, then enter all the words in the dictionary that express, explain, and reason;

Step 2: design basis layer extraction.

1) According to vocabulary that stored in the design basis vocabulary O, extract the sentences containing any of the words between the adjacent titles as the design basis candidate set L.

2) The design decision between the corresponding adjacent titles obtained by the design decision extraction algorithm and the previous step acquired design basis candidate sentences L as the nodes, constructing semantic syntactic graph G (B, X) [12]. The node B is defined as the vertex, and the connection weight is represented by the matrix X representing the semantic relevance of the node.  $x(b_i, b_j)$  represents the correlation between the two sentences  $b_i$  and  $b_j$ , among which  $x_{ii} = 0$ . The similarity matrix X is treated with symmetric normalization, as follows formula (3):

$$M = Q^{-1/2} X Q^{-1/2} \quad (3)$$

Q is a diagonal matrix,  $q_{ii} = \sum_{j=1,2,\dots,n} w_{ij}$ ; sentence score vector  $f = [f_0, f_1 \dots \dots f_n]^T$ , set initial value to  $f_i = \frac{1}{n}$ , define an n-dimensional vector  $y = [1, 1 \dots \dots 1]^T$ .

3) Calculate the weight between the design decision and the design basis. Cyclic calculation formula (4) until

the calculation results converge. Where r represents the rth calculation,  $\theta$  is the coefficient ( $0 \leq \theta \leq 1$ ).

$$f(r+1) = \theta M f(r) + (1 - \theta) \quad (4)$$

Step 3: design basis layer generation

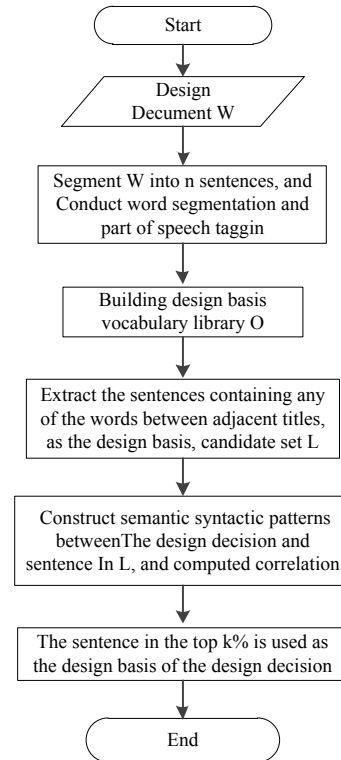


Fig. 4. Algorithm for design basis layer generation

Through step 2, the weight values between design decision and design basis are obtained. According to the weight value, the design basis which is most relevant to a design decision can be obtained. For each design decision take weight value in the former k% of the design basis, and build connection. Finally, the design intention-design decision-design basis hierarchy is formed.

## 4 Using design document as example

Fig. 5 shows the DR captured from a design document that focuses on welding robots. From the design intent layer, it indicates the motivations of a new design. These design intentions come from the title of the design document. In general, the design decision layer gives the knowledge of what to select to realize the design intent. The design basis layer shows the knowledge why to select the design decision to realize the design intent.

Fig. 6 shows our interface of DR knowledge extraction. The full name of this system is Cope DR Knowledge Model Editor (Cope-DRKM Editor). Cope-DRKM Editor is a system that extracts DR knowledge from design documents (word, PDF and other formats) and constructs a DR knowledge model. It has the functions of design intention extraction, design



decision extraction, design basis extraction, DR knowledge model construction, and so on. The system supports the knowledge processing and the efficient reuse of DR knowledge.

The first step is to fill in the information related to project, the second step is to input the design document, the third step is to adjust the design document title format, the fourth step extraction design intent, the fifth step is extraction design decision, the sixth step is extraction design basis, the last step is to modify the model. Because the wizard system, both the new and old users can easily complete the extraction task.

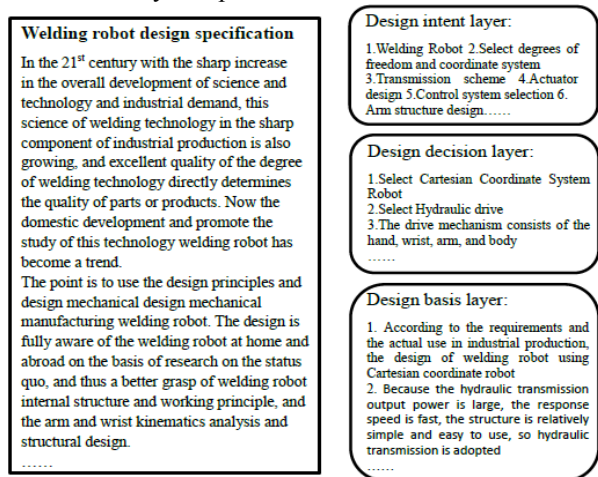


Fig. 5. An example of DR knowledge extracted

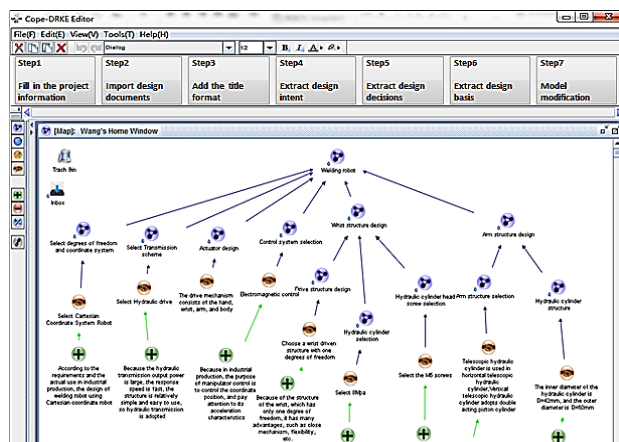


Fig. 6. Interface of DR knowledge extraction

## 5 Conclusion

This paper is based on the premise of not affecting the design process, extracting DR knowledge from a large number of historical design documents, and a structured and hierarchical DR knowledge model is used to express the DR knowledge. This method has promoted the acquisition efficiency and amount of DR, it lays a strong foundation for the DR knowledge base and the reuse of DR.

Because of computer recognition of natural language, resulting in model construction is not accurate enough, in the design decision layer and the design intent layer,

there may be some lack of decision and intent, or there is a node that is not a decision and intent. Further experimental studies will focus on the improvement of the model, so that the system can achieve automatic extraction and improve the extraction accuracy.

## Acknowledgements

This work has been supported by Project of National Science Foundation of China through approval No.51475027 and Project of National Key Technology R&D Program through approval No. 2015BAF18B01.

## References

1. J. Lee, K. Lai, What's in design rationale. *Human-Computer Interaction*, **6 (3)** :251—280 (2011)
2. J. Lee, Design rationale systems: understanding the issues. *IEEE Expert*, **12 (3)** :78-85 (1997)
3. W.C. Regli, X. Hu, M. Atwood, W. Sun, A survey of design rationale systems: approaches, representation, capture and retrieval, *Engineering with Computers*, **16 (3-4)** :209-235 (2000)
4. Y. Liu, Y. Liang, C.K. Kwong, W.B. Lee, A new design rationale representation model for rationale mining, *J. Comput. Inform. Sci. Eng.*, **10 (3)** :031009-1–031009-10 (2010)
5. Simon H. A, *The Sciences of the Artificial*, (M.I.T. Press, 1969)
6. McCall R. J., PHI: A Conceptual Foundation for Design Hypermedia, *Design Studies*, **12(1)**: 30-41 (1991)
7. MacLean A., Young R., Bellotti V., et al., Questions, Options, and Criteria: Elements of Design Space Analysis, *Human-Computer Interaction*, **6(3-4)**: 201-250 (1991)
8. Lee J., Decision Representation Language (DRL) and Its Support Environment, *Mit Artificial Intelligence Laboratory*, **36 (16)** :2307-2310 (1989)
9. Maher M., Tang H., Co-Evolution as a Computational and Cognitive Model of Design, *Research in Engineering Design*, **14(1)**: 47-64 (2003)
10. Ganeshan R., Garrett J., Finger S. A framework for representing design intent, *Design Studies*, **15(1)**: 59-84 (1994)
11. Liu J. H., Hu X. J. *Modeling the Evolving Design Rationale to Achieve a Shared Understanding*,(IEEE Press, 2012)
12. Y. Liu, Y. Liang, Learning the "Whys": Discovering design rationale using text mining - An algorithm perspective, *Computer-Aided Design*, **44 (10)** :916-930 (2012)