IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 20 November 2022, accepted 11 December 2022, date of publication 15 December 2022, date of current version 9 January 2023.

Digital Object Identifier 10.1109/ACCESS.2022.3229490

RESEARCH ARTICLE

A Novel Patent Knowledge Extraction Method for Innovative Design

GAOFENG YUE[1,2], JIHONG LIU[1], YONGZHU HOU[3], AND QIANG ZHANG[1]
[1]School of Mechanical Engineering and Automation, Beihang University, Beijing 100191, China
[2]China National Institute of Standardization, Beijing 100191, China
[3]Beijing Institute of Mechanical and Electrical Engineering, Beijing 100072, China

Corresponding author: Jihong Liu (ryukeiko@buaa.edu.cn)

This work was supported in part by the National Key Research and Development Program from the Ministry of Science and Technology of the Republic of China under Grant 2021YFB1716201, and in part by the Key Research and Development Program from the China National Institute of Standardization under Grant 532022Y-9424.

ABSTRACT As an important source of inspiration, the great number of patent documents provides designers with valuable knowledge of design rationale (DR), including *issues*, *intent*, *pros* and *cons* of the solutions. Researchers have carried out a number of data analysis studies based on patent information, which is now a new discipline called Patinformatics, including the analysis of patent information from a macro perspective and the identification and extraction of patent knowledge from a micro perspective. If DR knowledge could be extracted automatically from the patent documents and provided to designers as a source of inspiration, it would greatly promote innovative design, and at the same time promote the reuse of patent documents and the wide application of DR theory, which can be like killing three birds with one stone. To address this issue, this study proposes an improved lexical-syntactic pattern method for DR centric patent knowledge extraction, including DR Vector Space model (DRVS), DRV Trigger Word (DRV-TW), Design Rationale Vector (DRV), DR credibility (*DRC*) and others, and DRV based knowledge extraction algorithms. Knowledge extraction experiments were conducted on 1491 patent documents to verify the feasibility and performance of the method. In addition, two other sets of comparative experiments were conducted using the FastText and BERT machine learning methods, and the results further confirmed the reliability of the proposed method for low-resource corpus.

INDEX TERMS Patent analysis, design rationale, knowledge extraction, design knowledge network.

I. INTRODUCTION

Innovative design is a typical knowledge-intensive activity. Designers need to retrieve a large amount of existing design knowledge and information, find other knowledge and solutions that can be used for reference, and provide valuable inspiration, ideas or decision support [1], [2], [3], [4]. The openness of innovative design is a common feature of both art and engineering. The novelty of each candidate design is primarily dependent on the process of inspiration and on the way information is integrated during the generation of new solutions [2]. Sources of inspiration can help designers define context, trigger ideas, and build a designer's mental representation during the design thinking process, which is becoming a central point for industries to seek innovative solutions to problems [2], [5]. DR is important design knowledge about the design process, which includes *issues*, *intents*, *alternatives*, *pros* and *cons* of the design. DR plays a very important role in engineering design, such as facilitating collaborative design [6], [7], [8], [9], [10], [11], helping designers with personal knowledge management, assisting in relevant education and training [12], implementing root cause analysis to track design failures [13], [14]. If DR could be automatically extracted from patent documents and formally stored, it would be of great value to designers, because the extracted DR can inspire innovative designs.

Patented technologies represent newer technologies and research results in various fields due to its unique

The associate editor coordinating the review of this manuscript and approving it for publication was Agostino Forestiero.

characteristics, such as novelty and utility. A large number of published patents from all over the world is of great value to the design of innovative products [3], [4], [15]. With the wide application of patent information in different fields, Patinformatics has been proposed [16] as a new discipline, covering macro-level data analysis and micro-level patent information extraction. Macro trend analysis includes analysis and prediction of future technology trends [17], human resource development [18], [19], patent infringement analysis [20], technology intelligence tool [21], technology concentration and specialization analysis between countries [22] and others. Microscopic patent analysis is mainly about the automatic knowledge extraction from patent documents.

Patent documents have multiple writing styles, addressing various issues and purposes, different language habits and professional backgrounds, which make it difficult to read and comprehend, and difficult to directly serve as a reference material for designers [23], [24]. With the help of natural language processing, machine learning, information retrieval, citation analysis and other methods, researchers try to extract pertinent knowledge from patent documents to gain insight into design problems [4], [25]. Patent analysis is used to capture scientific effect knowledge [26], extract function-behaviour-state (FBS) knowledge [27], and extract functional knowledge [3]. Trappey et al. [28] proposed a method to automatically generate summary reports in a given domain using artificial intelligence, natural language processing, deep learning techniques and machine learning algorithms. Liu et al. [29] proposed a representation model for rationale information discovery from design archival documents. Based on the ISAL model, Liang et al. [30] proposed a method to extract DR from patents, including *issues*, *solutions* and *artifacts* information. In recent years, Sarica et al. [25], Sun et al. [31], Zuo et al. [32] and others conducted research on acquiring and constructing engineering knowledge graphs (KG) from patents.

Only the ISAL model and the related knowledge extraction research is about DR knowledge. However, from the perspective of providing a source of inspiration for designers, the methods still have some deficiencies. The extracted rationale information is relatively coarse, and it includes merely the information about *issues*, *solutions* and *artifacts*, and some most important rationale information is ignored, such as *issues*, *intents*, *pros* and *cons*, and *alternatives*. Besides, it is mainly paragraph-level knowledge, rather than words, phrases or sentences that can provide designers with intuitive inspiration.

To address the above issues, a novel patent analysis method for automatic DR extraction and knowledge graph construction is proposed. The main contributions of this paper can be summarized as follows:

- A DRVS model is proposed based on the vector space model and lexical syntactic pattern. It can be used for DR identification and extraction for low resource corpus, which integrates a variety of Natural Language Processing (NLP) tasks such as sentiment analysis, named entity recognition, and relationship extraction.
- The method supports the automatic construction of design knowledge graphs, which can provide creative stimuli for designers, and can further be used as an intelligent tool for macro analysis, such as human resource management, design history analysis and others.
- This method has the characteristics of high reliability, ease of use, and low resources, and is especially suitable for discovery of engineering knowledge from patent documents.

The rest of this paper is organized as follows. In Section II, we review relevant work on DR and design knowledge extraction from documents, as well as NLP methods. Section III introduces the proposed model and methodology to address knowledge extraction. Section IV presents an empirical study on knowledge extraction and DKN construction. Finally, Section V gives conclusions and future work to be done.

II. RELATED WORK

A. DESIGN RATIONALE AND DESIGN KNOWLEDGE EXTRACTION FROM DOCUMENTS

With the rapid growth of digital libraries, the large amount of scholarly data poses increasing challenges for researchers and publishers to analyze scholarly information. M. et al. [33] proposed a graph-based approach consisting of nodes and edges to deal with the big data problems of scholarly literature, including selecting the right reviewers for submitted papers, finding quality impact factors, and ranking journals and researchers.

Patent knowledge extraction based on TRIZ theory for engineering design is an important research branch of patent analysis. Souili et al. [34] proposed an IDM (Inventive Design Method) ontology based on TRIZ theory, combined with the lexical syntactic pattern (LSP) approach, to extract IDM knowledge from patent databases, including problems, partial solutions, and contradictions. The extraction process consists of four main steps: selecting the relevant text areas; segmentation of patent documents at the paragraph level; using finite state automata to match and label IDM-related knowledge; knowledge extraction and problem graph generation.

Based on TRIZ theory, effect knowledge is about using scientific laws to achieve the desired product functions, which is of great value to designers. Aiming to assist in achieving high level product innovation, Liu et al. [26] proposed a method that combines syntactic analysis, WordNet and word vector technologies for extracting the required effect knowledge from International Patent Classification (IPC) texts.

To extract the motivating problem (contradiction knowledge) and achieve a rapid understanding of patent content, Guarino et al. [35] proposed a patent analysis method based on a combination of sentence-level and word-level deep neural networks.

Valverde et al. [4] proposed a method of patent knowledge extraction to inspire designers with ideas and analogous solutions during the problem-solving phase. The knowledge extracted is based on TRIZ theory, including problems, functions, physical effects, technology evolution trends and others. All knowledge extraction processes in the study were carried out manually.

Sun et al. [31] proposed a metamodel to build the KG of design objects, and a knowledge extraction method based on dependency parsing analysis to help non-Chinese designers and scholars to extract valuable design knowledge from Chinese language patents.

To discovery the rationale information from design archival documents, Liu et al. [29] proposed an ISAL model that includes *issues*, *solutions*, and *artifacts*. On this basis, Liang et al. [30] proposed a method for extracting DR paragraphs from patent documents, including artifact information, issue summarization and solution–reason pairs. The method is based on the analysis of term frequencies, language patterns and manifold ranking.

Lester et al. [36] conducted a rationale extraction study on bug reports of Chrome web browser, which included *decisions*, *alternatives*, *answers*, *arguments*, *assumptions*, *procedures*, *questions*, and *requirements*. They studied two evolutionary algorithms to optimize feature selection to improve knowledge extraction performance based on NLTK's Naive Bayes classifier.

Kurtanović and Maalej [37] studied online reviews in the Amazon Store on how users argue and justify their software selections, including *issues*, *alternatives*, *criteria*, *decisions* and *justifications*. Several extraction algorithms were used, including Naive Bayes, Decision Tree, Support Vector Classifier, Logistic Regression, Gaussian Process Classifier, Random Forest, and Multilayer Perceptron Classifier.

Although so many studies focus on the extraction of design knowledge using NLP technology, they mainly focus on extracting scientific effect knowledge, design object knowledge, and function knowledge based on TRIZ. Few studies have focused on extracting DR knowledge from patent documents to provide designers with creative inspiration.

B. NATURAL LANGUAGE PROCESSING

Sentiment analysis is an NLP method that analyzes people's opinions or sentiments in written texts [38]. Sentiment analysis is used to help extract location information from social networks [39], evaluate global supply chain risk [40], and identify customer satisfaction scores from customer review data [41]. Dictionary-based sentiment analysis, a sentiment dictionary needs to be prepared in advance, including positive words and negative words. Then, traverse all the words in the target sentence and count the number of positive words and negative words. The dictionary-based sentiment analysis method is a mature method with high precision. The method requires a sentimental dictionary in advance, including positive and negative words; Sentiment analysis is then performed by iterating through all the words in the target sentence and counting the number of positive and negative words.

Lexical syntactic pattern for relation extraction. Relation Extraction from unstructured text is the core task of KG construction, which has received extensive attention in recent years. There are generally three approaches of relationship extraction: (1) co-occurrence based, (2) machine-learning based [42], [43] and (3) rule-based. The co-occurrence-based approach [44], [45] is built on the assumption that entities occurring frequently together within a document and have higher chances of being related [43]. However, this is often not the case, and the precision of relation extraction is low [43]. Rule-based approaches use a set of rules or patterns, defined manually or automatically, to extract relations [43], [46], [47]. There are generally three machine learning paradigms for relation extraction [48], which include: (a) supervised approaches focusing on hand-labelled datasets, (b) unsupervised approaches targeting large amounts of text, and (c) bootstrap learning method that starts with small seed instances to iteratively learn patterns and entity pairs. Supervised learning paradigm typically relies upon large sets of labeled data, and unfortunately it is not readily available in real applications [49], [50]. Unsupervised learning paradigm works only with the input data without target variables. Therefore, there is no teacher to correct the model, as there is in supervised learning [51]. Rule-based methods, also called knowledge-based methods, use patterns and rules crafted by human experts for relation extraction from domain text. Pattern-based information extraction methods have a long history as a successful approach for domain-specific relation extraction [52]. Lexical-syntactic pattern was first proposed by Hearst [53] to extract hyponymy lexical relations from unrestricted text. Subsequently, lexical-syntactic pattern method was used in many domains [54], [55], [56]. Besides, lexical-syntactic pattern can be used for text categorization [57]. However, these patterns suffer from recall problem and precision problem. Zhou et al. [58] introduced the concept of trigger words, which activate patterns of specific relations and act as conceptual anchor points of patterns. Activation force-based trigger word mining was proposed to improve the performance of the relation extraction [59], [60]. Dependency patterns [61] have a better performance for more informative relation extraction [59]. The most popular dependency pattern method may be the shortest dependency path, which is used for a number of application domains [62], [63], [64]. FrameNet [65] provide annotations in terms of semantic frames comprising frame elements. Mandya et al. [52] proposed frame-based semantic patterns for relation extraction.

Knowledge fusion is an effective solution to the problem of acquiring knowledge from different sources [66]. Addressing the issue of multi-dimensional, heterogeneous, and time series on the data collection in the manufacturing process, Liu et al. [67] used knowledge fusion to disambiguate, integrate and reason the knowledge, and get high-quality knowledge after eliminating a series of redundant and erroneous information. Knowledge fusion process includes alignment or cognition matching of entities, relations, and attributes, which

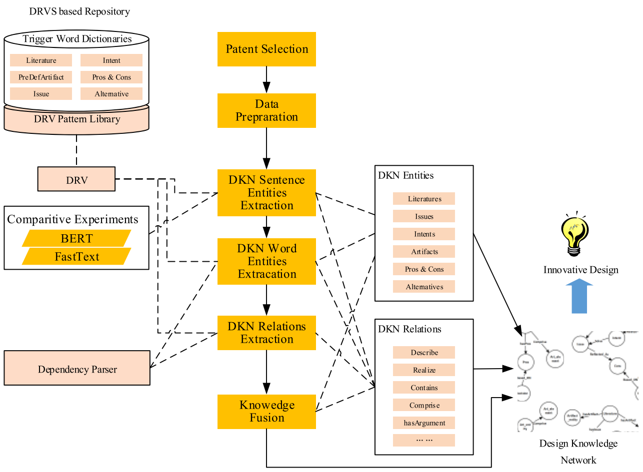


FIGURE 1. The framework of proposed method.

is one of the fundamental techniques for building KGs [67], [68], [69].

Vector Space Model. Salton et al. [70] proposed a vector space model, in which keyword or index terms are viewed as basic vectors in a linear vector space, and each document is represented as a vector in such a space. Each document Di can be represented by a t-dimensional vector, as shown in Expression (1):

$$\mathrm{Di} = (\mathrm{d_{i1}}, \mathrm{d_{i2}}, \ldots, \mathrm{d_{it}}) \tag{1}$$

where the $\mathrm{d_{ij}}$ representing the weight of the $\mathrm{j_{th}}$ term. In other words, each document space can be regarded as composed of t text vectors.

III. PROPOSED METHOD

A. OVERVIEW

To provide creative inspiration to designers, it is first necessary to build a DR-centric Design Knowledge Network (DKN). This requires extracting DKN entities and relations from patent documents that are critical to designers. Due to the special requirements of the patent knowledge extraction, this paper proposes a knowledge extraction method based on DRVS model. The method relies on a pre-established repository, including feature word dictionaries which can be seen as a DRVS in the real world, and a DRV pattern library. As shown in Fig. 1, the method includes the following steps:

Patent selection. Patent documents, as the source of knowledge, can be collected from the official website of the Patent Office through keywords retrieval or patent classifications.

Data preparation. This requires removing noisy data from patent data and separating patent documents into lists of sentences.

DKN sentence entity extraction. The DRV method is used to identify and extract DKN sentences such as *issues*, *alternatives*, *artifacts*, *intent*, *arguments*, and others. In addition, two machine-learning methods, BERT and FastText, are used as comparative experiments to extract DKN sentences.

DKN word entity extraction. Some DKN entities, such as *artifacts*, *pros* and cons, appear in the form of feature words; Design intent exists in the form of a phrase or purpose clause. Combined with the Dependency Parser (DP), the entities in form of words, phrases, clauses in the patent document are extracted.

DKN relation extraction. Combined with DRV, DP and patent document structure, the related relations are extracted.

Knowledge fusion. Entities and relations are merged to eliminate redundant knowledge.

B. DESIGN KNOWLEDGE NETWORK

From an engineering designer's point of view, DR provides valuable design inspiration for innovative designs. To build a DR-centric design knowledge network, the following entities and relations need to be extracted from patent documents:

Literature entity. Literature refers to patent documents and related metadata in the field of engineering design. From a linguistic point of view, technical literature includes documents, paragraphs, sentences, phrases and words.

Artifact entity. Artifacts, also called design objects in this study, can be anything that aims to achieve a goal, purpose, or function that satisfies human desires [71], and is the material basis for the realization of design intent. As used herein, artifacts include machinery, equipment, devices, computer programs, processes/methods, chemical components and other man-made objects.

Rationale entity. DR explains how and why products are designed that way [71]. In this paper, rationale entities include the following:

- *Issue*: A brief description of the problems or requirements of existing artifacts.
- *Intent*: The design goal that the designer wants to achieve through the artifact. Design intent can be functions, behaviors or performance of an artifact. Design objects and their components can generally perform certain operations to obtain certain functions, to achieve specific design tasks or goals. In addition, the design object reflects certain performance, such as reliability, safety, economy and efficiency, when fulfilling a function.
- *Argument*: A description of the advantages and disadvantages of the artifact, supportive or oppositive comments, opinions or sentiment analysis from the designer or interested parties. It includes: *Cons*, including deficiencies, shortcomings, and other negative descriptions of *alternatives* or the status quo; *Pros*, which are used to express positive information such as the excellent functions, reliable performance, and a wide range of application prospects.
- *Alternative*: Alternative artifacts, design options for reference, or related solutions quoted in the document.

Relations. We use the expression rules prescribed by Neo4j Cypher to represent nodes and edges: "()" denotes entities or nodes, and "-[]->" denotes logical relationships or edges with directions.

- ***Is_a*** relationship can be applied to represent the internal inheritance relationship between nodes, as shown in Expression (2).

$$(subClass) - [Is\_a] -> (Class) \tag{2}$$

- ***Contain*** can be applied to aggregation relationship, as shown in Expression (3) and (4).

$$(literature) - [hasDRSent] -> (Sentence) \tag{3}$$
$$(Sentence) - [hasDRWord] -> (Word) \tag{4}$$

- ***Describe*** relationship can be used for *literature* entities describe *artifact* entities, as shown in Expression (5).

$$(literature)\ \text{-}[Describe] -> (artifact) \tag{5}$$

- *Artifact* entities ***Realize*** a certain *intent*, as shown in Expression (6).

$$(artifact) - [Realize] -> (intent) \tag{6}$$

- ***Be_structured*** relationship. An *artifact_component* is composed of *artifact_parts*, which can be embodied as *Comprise* relationship, as shown in Expression (7).

$$(artifact_{c}omponent) = (artifact_{p}art) - [Be_{S}tructured] -> (artifact_{p}art) \tag{7}$$

- ***HasOpinion*** relationship. Designers hold a *negative* or *positive* opinion through an *argument* sentence, as shown in Expression (8).

$$(argument)\ \text{-}[hasOpinion\ ] -> (position) \tag{8}$$

- ***There_Exist***(or ***hasIssue***) a design problem or design requirement in the literature, as shown in Expression (9).

$$() - [There_{Exist}] -> (issue) \tag{9}$$

- ***HasAlternative*** represents that the artifact has one or more *alternatives*, as shown in Expression (10).

$$(artifact)\text{-}[has\ Alternative] -> (alternatives) \tag{10}$$

C. FRAMEWORK OF THE DRV-BASED METHOD

Compared with the Automatic Content Extraction (ACE) task and the Open Information Extraction (OIE) task, DR knowledge extraction from patent documents has the following characteristics:

- The precision and reliability of the extracted information are more important than the recall. If the extracted knowledge is correct, even if only part of DR is extracted, it is still beneficial for designers due to the huge number

$$DRVS = (f_1, f_2, f_3, \ldots, f_n) \tag{11}$$
$$P: S_{d\,r} \leftarrow DRVS \tag{12}$$

of available patent documents.
- The grammatical structure of patent documents is relatively regular. Patent documents are usually drafted by professionals in accordance with strict patent drafting rules. This is more suitable for syntactic pattern analysis as a knowledge extraction method.
- The entities, relationships, and scenarios in this task are limited to the engineering design domain, which determines that it is not a generic NLP task. In addition, there is no publicly available corpus for this study, and there is no uniform knowledge extraction target.

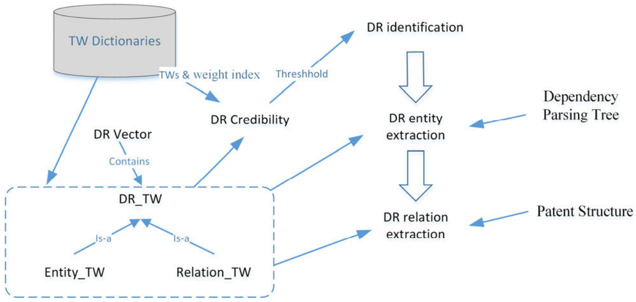


FIGURE 2. DRV-based knowledge extraction.

To address these issues, this paper proposes a dictionary-based lexical syntactic pattern, the DRV method. As shown in Fig. 2, the core parts of the proposed methods are described as follows:

- DR Vector is the basis for DR identification, related entities and relations extraction.
- DRV-TW based on activation force is the basic element of DRV.
- DR Credibility based on DRV-TWs is the weight index to identify whether the target sentence contains DR knowledge.
- DRV-based algorithms, combing the DP and patent structure, are used for DR identification and extraction.

D. DESIGN RATIONALE VECTOR SPACE MODEL

A sentence is the unit that constitutes a patent document; Sentences are made up of words according to a certain syntactic pattern. Most DR sentences in the patent have typical syntactic patterns, as shown in Table 1. Based on the Vector Space Model [70] and Lexical-Syntactic Patterns, we proposed the concept of Design Rationale Vector Space, which is composed of DR feature word lists, as shown in Expression (11). where the f_n is a list of feature words of a particular type. A DR sentence (S_{dr}) of patent documents can be expressed in DRVS according to a certain syntactic pattern P, as shown in Expression (12)

Trigger Words. The most important feature words, which can be called trigger words (TW), determine the core meaning of the whole sentence. Similar to Frame Element

TABLE 1. DRV instances.

DRV type	DR Instance
Lit_Des_Art_Rea_Int DRV	Embodiments disclosed herein [*literature*, I_{lib} =0.9] **provide** [*Describe*, I_{des}=0.9] systems [*preDefinedArtifact*, I_{art}=0.9] and methods[*preDefinedArtifact*, I_{art}=0.9] **for** [*Realize*, I_{rea}=0.9] obstacle detection and state information determination [*intent*]
Lit_Art_Rea_Int_Des DRV	In one embodiment[*literature*, I_{lib} =0.9], an infrared (IR) imaging system [*preDefinedArtifact*, I_{art}=0.9] **for** [*Realize*, I_{rea}=0.9] determining a concentration of a target species in an object [*intent*] **is disclosed** [*Describe*, I_{des}=0.9]
Art_Comp_ArtEle_Rea_Int DRV	A cargo transport system [*preDefinedArtifact*, I_{art}=0.9] ***includes*** [*Comprise*, I_{com}=0.9] a UAV[*artifact_element*] and a *vehicle* [*artifact_element*] **for** [*Realize*, I_{rea}=0.9] sending and receiving the UAV [*intent*].
ThereExist_Issue DRV	From the above, it is evident that ***there remains [There_Exist, I_{thr}=0.9]*** a need [*Issue_word,* I_{iss}**=0.9]** in the industry for more efficient demining techniques that do not give rise to at least some of the issues described above.
Argument	However, with the connecting member having the elastic snap structure[*artifact*], **poor engagement [***cons*, $I_{con=}$**0.9]** may occur when the connecting member is snap-fitted to the wire connector due to **low manufacture precision [***cons*, $I_{con=}$**0.9]** of the connecting member; and shake generated by the aerial vehicle during flight may also lead to **poor contact [***cons*, $I_{con=}$**0.9]** of the snap structure, thereby adversely affecting the supply of power to the aerial vehicle; moreover, the elastic snap may be prone to **deformation [***cons*,$I_{con=}$**0.8]** and **damage [***cons*, $I_{con=}$**0.9]** after long-term use, which may **shorten [***cons*, $I_{con=}$**0.6]** the service life thereof.
Alternative	***According to Specification of*** [*hasAlternative*, I_{ref}=0.9] **U.S. Pat. No. 6,540,179** [*alternative*, I_{alt}=0.9], landing and take-off of an unmanned aircraft presents problems in providing necessary communication links between ground controllers and the unmanned aircraft.
Composite DRV： *Lit_Des_Art_Rea_Int* & *Art_Comp_ArtEle_Rea_Int*	In one aspect[*literature*, I_{lib} =0.9], the present invention[*literature*, I_{lib} =0.9] **provides** [*Describe*, I_{des}=0.9] a system [*preDefinedArtifact*, I_{art}=0.9] ***for*** [*Realize*, I_{rea}=0.9] maintaining equipment within a predetermined area [*intent*], ***including*** [*Comprise*, I_{com}=0.9] a first unmanned vehicle [*artifact_element*] ***configured to*** [*Realize*, I_{rea}=0.9] perform a diagnostic evaluation of the equipment [*intent*], a second unmanned vehicle[*artifact_element*] ***configured to*** [*Realize*, I_{rea}=0.9] perform a maintenance operation [*intent*], and a third unmanned vehicle [*artifact_element*] ***configured to*** [*Realize*, I_{rea}=0.9] perform a safety operation [*intent*].

in FrameNet, TW can be grammatically functional words, which may be a single word, or a phrase, or a regular expression representing a functional entity. For example, *alternative* TW can be represented as the following regular expressions:

"US([0-9]{4,})|U.S. Pat. No. [0-9],"

"US ([0-9]{4,})"

"U.S. Pat. No."

"US Patent Application ([0-9]{4,})"

TWs include the following types: *positive* words or *pros*; *negative* words or *cons*; *literature* words; *predefined_artifact* words; *issue* words; *alternative* words; *Describe* words; *Realize* words; *There_Exist* words; *Be_Structured* words. These TWs are generally generic and frequently used in most patent documents, which is a type of controlled word from the TW dictionaries.

As shown in expression (13) ~ (17), the TWs in a DR sentence constitutes a DRV according to a certain syntactic pattern. The related instances are shown in Table 1.

$$\text{(literature)-[Describe]}{-}{>}\text{ (artifact)-[Realize]}{-}{>}\text{ (intent)} \tag{13}$$

$$(Artifact_{e}ntity) - [Comprise]{-}{>} (artifact_{e}lement) - \text{[Realize]- > (intent)} \tag{14}$$

$$() - [There_{\text{Exist}}]{-}{>} \text{(issue)} \tag{15}$$

$$\text{(artifact)} - \text{[hasAlternative]}{-}{>} \text{(alternative)} \tag{16}$$

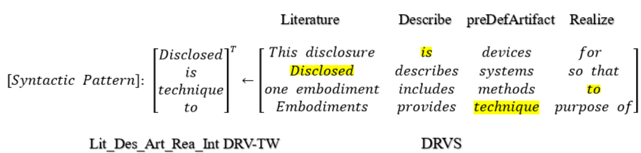


FIGURE 3. Illustrative diagram of DRV transferred from DRVS according to the syntactic pattern rules.

$$\text{(artifact)} - \text{[hasArguments]}{-}{>} \text{(Pros\&Cons)} \tag{17}$$

To reduce complexity of the study, the constituent element of DRVS is restricted to TWs instead of general feature words. A DRVS is a matrix space consisting of DRV-TWs, and is a database that consists of lists of DRV-TWs and their weight index in real world. The TWs in a DRV can be viewed as: a vector represented with the DRVS matrix through some transformation, as shown in Fig. 3; a point in the DRVS coordinate system, as shown in Fig. 4.

DRV-TW can be used to represent key semantics of DRVs. DRV methods rely on the domain dictionary established in advance, including the TW dictionaries and the DRV pattern library, as shown in Fig. 1. The product ontology that defines the property hierarchy of a product or part family [73], [74] is a database of design objects and can serve as the basis to create a dictionary of *predefined artifacts*.

The *intent* has no fixed feature words, which is a kind of design objective information from a pragmatic point of view.

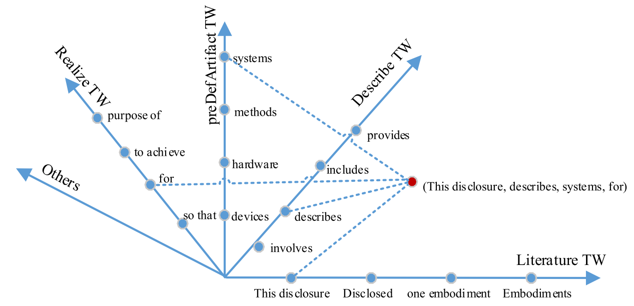


FIGURE 4. Illustrative diagram of a DRV in DRVS.

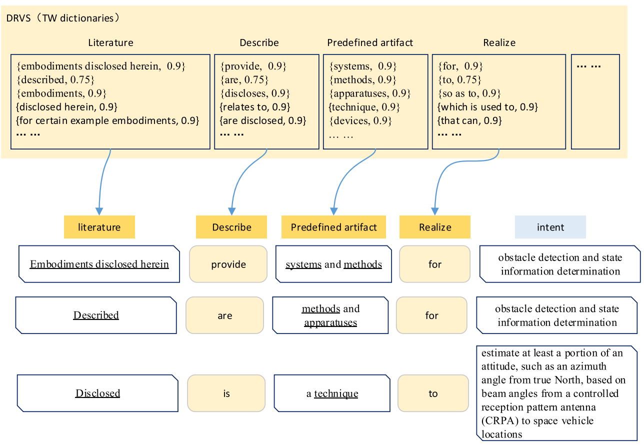


FIGURE 5. Intent as pragmatic information extracted from DR sentences.

As shown in Fig. 5, based on the DRV shown in Expression (13), and the TW dictionaries, DR sentences can be identified from the target sentence list, and the relevant DR entities, such as *literature*, *artifacts*, *intents*, can be extracted with the aid of the NLP syntactic parsers. Table 1 is some instance sentences, and related DRV_TWs and their weight indexes. The word in square brackets [] represents the type of DRV-TW. Composite DRV is composed of several simple DRVs. The last two rows of the table are simple DRVs: *Lit_Des_Art_Rea_Int* and *Art_Comp_ArtEle_Rea_Int*.

A DRV-TW list representing the relevant DRV pattern can be used to identify the corresponding DR sentence with its entities and relations. At the same time, these trigger words can be used to represent the key meaning of the DR sentence. The relevant DRV-TW list example is as follows:

[“*literature*”, “*preDefinedArtifact*”, “*Realize*”]
[“*Describe*”, “*preDefinedArtifact*”, “*Realize*”]
[“*preDefinedArtifact*”, “*Comprise*”, “*Realize*”]
[“*There_Exist*”,“*issue*”]
[“*artifact*”, “*pros*”, “*cons*”]
[“*Comprise*”]
[“*HasAlternative*”, “*alternative*”]

Each TW has a specific weight index for the DRV, and these weight indexes indicate the importance of the TW for the expressed semantics. DR Credibility (*DRC*) is an index used to determine whether a sentence matches the corresponding DRV pattern. The calculation formula of *DRC* is shown in Expression (18):

$$DRC = Max\ (I_{lib}, I_{des}) * I_{art} * I_{rea} \tag{18}$$

where I_{lib}, I_{des}, I_{art} and I_{rea} the weight indexes for literature TW, *Describe* TW, *preDefinedArtifact* TW, and *Realize* TW. For the TW of *preDefinedArtifact*, the weight index(I_{art}) can increase as the number of occurrences increases, but can never be greater than 1. The formula of I_{art} is shown in Expression (19):

$$I_{art} = \frac{2}{\pi} \times arctan(\sum\nolimits_{k=1}^{n} I_{obj.k}) \tag{19}$$

where $I_{obj.k}$ the k_{th} weight index of the *preDefinedArtifact* TW. The reference *DRC* formulas of other DRVs are listed in Table 2∼5.

TABLE 2. Algorithm 1: extraction of ***issues***.

Input	1. Patent documents; 2. *Issue_vectors*, for example: -*[There_Exist]->(issue)* 3. *Issue_vector_TW* lists, such as [*There_Exist* TW, *issue TW*]; 4. TW dictionary containing TWs and corresponding weight indexes; 5. T_{issue}, the threshold of DRC_{issue} for identification of *issue* sentences.
Procedure	1. Preprocess text to remove noisy data; 2. Segment the main text into sentence lists; 3. Read the corresponding TWs from the dictionary and their weight indexes according to the *issue_TW* list; 4. Determine whether the target sentence contains the *issue_TW* in *feature_list* through the regular operation; if it does, record the *issue_TW* and their weight indexes; 5. Calculate the *DRC*: $DRC_{issue} = \mathrm{Max}(I_{thr}) * I_{iss}$ 6. Comparing DRC_{issue} with the threshold T_{issue}: If DRC_{issue} > T_{issue}: Then extract the sentence entities and relations.
Output	Extracted issue sentences and *hasIssue* relation.

E. DR IDENTIFICATION AND EXTRACTION

Fig. 6 shows a flowchart of DR identification and extraction, with patent documents input and DR entities and relationships extracted. The algorithm mainly includes the following steps: pre-initialize the DRV_TW list, which is from different DRV syntactic patterns; traverse and query whether the TWs in the DRV_TW list are included in both the TW dictionary and the word list of the tokenized target sentence; If the target sentence contains all TWs in the DRV_TW list, the corresponding weight indexes of TW is read from the dictionary; The DRC can be calculated from the relevant DRC formula and compared to the threshold, so that DR sentences can be accurately identified.

TABLE 3. Algorithm 2: extraction of *artifacts* and intents.

Input	1. Patent documents; 2. *Intent_vectors*, for example: *(literature)--[Describe]-->(artifact)-[Realize]-->(intent)* *(Artifact_entity)--[Comprise]-->(artifact_element)-[Realize]-->(intent)* 3. *Intent_vector_TW* lists, such as ["*literature*", "*preDefinedArtifact*", "*Realize*"], ["*preDefinedArtifact*", "*Comprise*", "*Realize*"] 4. TW dictionary based on *intent_vector*, including related TW and their weight indexes, such as *literature*, *Describe*, *preDefinedArtifact*, *Realize*; 5. T_{intent}, the threshold of DRC_{intent} for identification of *intent* sentences.
Procedure	1. Preprocess text to remove noisy data; 2. Segment the main text into sentence lists; 3. Read the corresponding type of TWs from the dictionary and their weight indexes according to the *Intent_vector_TW* list; 4. Determine whether the target sentence contains the TWs in *feature_list* through the regular operation; if it does, record the corresponding TW and their weight indexes; 5. Calculate the *DRC*: $DRC_{intent} = \text{Max}(I_{lib}, I_{des}) * I_{art} * I_{rea}$ 6. Comparing DRC_{intent} with the threshold T_{intent}: If $DRC_{intent} > T_{intent}$: Then extract the sentence.
	7. If the target sentence contains design intent, extract *intent* through syntactic parser according to the *Intent_vector*; 8. Extract the related TW contain in the target sentence as the DR, including *literature*, *artifacts_components* and related *artifact_parts*, *Describe*, *Realize*, *hasArtifact*, *hasIntent* and other relations.
Output	1. Extracted *intent* sentences; 2. Extracted entities, including literature, *artifact_components* and related *artifact_parts*, *intent*; 3. Extracted relations, including *hasIntent*, *hasArtifact*, *Describe*, *Comprise* and *Realize* relations.

TABLE 4. Algorithm 3: extraction of arguments.

Input	1. Patent documents; 2. TW dictionary that contains negative TWs and positives TWs (including TWs and their weight indexes); 3. $T_{argument}$, the threshold of $DRC_{argument}$ for identification of *argument* sentence, including *pros* and *cons* sentence.
Procedure	1. Preprocess text to remove noisy data; 2. Segment the main text into sentence lists; 3. Read the *pros* and *cons* TWs from the dictionary and their weight indexes; 4. Determine whether the target sentence contains the TWs in *feature_list* through the regular operation; if it does, record the corresponding TW and their weight indexes; 5. Calculate the *DRC*: $DRC_{argument} = \sum_{k=1}^{n} I_{arg\_i}$ 6. Comparing $DRC_{argument}$ with the threshold $T_{argument}$, including T_{pros} and T_{cons}: If $DRC_{argument} > T_{pros}$: Then extract the *pros* sentence. Elif $DRC_{argument} < T_{cons}$: Then extract the *cons* sentence. 7. If the target sentence contains design argument, extract argument sentence; 8. Extract the related entities and relations contain in the target sentence as the DR, including *pros* and *cons*, and *hasArgument*, *hasOpinion* relations.
Output	1. Extracted argument sentences; 2. Extracted argument TWs, including *pros* and *cons*; 3. Extracted *hasArgument*, *hasOpinion* relations.

The algorithms for DR identification and extraction based on DRV-TW are shown in Table 2∼5.

After DR identification is completed, the relevant entities and relations can be further extracted from the DR sentence based on DRV patterns, including *literature*, *predefined artifacts* and *arguments* (*pros and cons*). Combined with NLP syntactic parser, DR entities can be extracted, including *intents*, *artifact_components*, *artifact_parts*. For *issues* and *alternatives*, use the entire sentence as the DR entity.

TWs as DR entities. According to the DR identification method mentioned above, if the target sentence is identified as containing DR information, this means that the TWs in the DRV are feature words of the target sentence and can be extracted as DR information. As shown in Fig. 7.

Syntactic parser aided DR entities extraction. For *intents*, *alternatives* and *artifacts*, including *artifact_components* and *artifact_parts*, other than the *pre-defined artifacts* in the dictionary, StanfordNLP and NLTK's NLP toolkit, such as Part-Of-Speech, tag(POS-tag), Named Entity Recognition (NER), Syntactic Parse and others, can be used for DR extraction, as shown in Fig. 8.

Extraction of*issues*or*alternatives*. For *issues* and *alternatives*, the semantics of TW alone are not rich enough to provide creative inspiration for designers. Therefore, the entire sentence needs to be used as the DR entity.

Only by realizing the automatic identification and extraction of relationships between nodes can the automatic construction of DKN be realized. Therefore, for different DR relationships, we propose three extraction methods.

DRV-based relation extraction. In this study, we hypothesize that if a sentence in a patent document contains a specific DRV pattern, then it contains the corresponding type

TABLE 5. Algorithm 4: extraction of *alternatives*.

Input	1. Patent documents; 2. Alternative_vector used to represent *alternative* sentence, for example: *(artifact)--[HasAlternative]-->(alternative)* 3. *Alternative_vector_TW* lists, such as ["*hasAlternative*", "*alternative*"]; 4. TW dictionary based on alternative_vector (including related TW and their weight indexes); 5. $T_{alternative}$, the threshold of $DRC_{alternative}$ for indentification of *alternative* sentences.
Procedure	1. Preprocess text to remove noisy data; 2. Segment the main text into sentence lists; 3. Read the corresponding type of TWs from the dictionary and their weight indexes according to the *Alternative_vector_TW* list; 4. Determine whether the target sentence contains the TWs in *feature_list* through the regular operation; if it does, record the corresponding TW and their weight indexes; 5. Calculate the *DRC*: $DRC_{alternatives} = I_{hasAlt} \times 2/\pi \times \arctan(\sum_{k=0}^{n}(I_{art_k}))$ 6. Comparing $DRC_{alternative}$ with the threshold $T_{alternative}$: If $DRC_{alternative} > T_{alternative}$: Then extract the sentence and relation.
Output	Extracted *alternatives* and *hasAlternative* relation.

of DR. For some DR relations, such as *Describe*, *Realize*, *Comprise*, *hasAlternative*, and *hasArgument* and others, specific DR sentences can be identified by a DRV-based syntactic document parser, thereby extracting specific DR entities and relations. For example, the relation between *Describe* and *Realize* is extracted through the Expression (15).

Linguistic structure-based relation extraction. The linguistic structure of patent literature is used to assist the extraction of DR relations. The different parts of a patent are called Basic Document Units (BDUs) in this study, such as "Abstract", "Introduction", "Background", "Abstract". There is a consistent DR relationship within the same BDU or between different BDUs in the same patent document. For example, the *intent* described in the "Abstract" should *solve* the *issue* described in the "Background" of the same patent document. In addition, there are several levels in patent documents, including documents, sentences and words, depending on the granularity of the knowledge contained in the document. The Contain relationship in a document can be extracted based on this hierarchy. DR from different patents can be interconnected through the citation relationship, forming a systematic knowledge network. Fig. 9 shows an example of the two-tier citation relationship of patent literature. In USPTO patents, there are some contents identified by "References Cited", "Other References", "CROSS-REFERENCE TO RELATED APPLICATIONS",

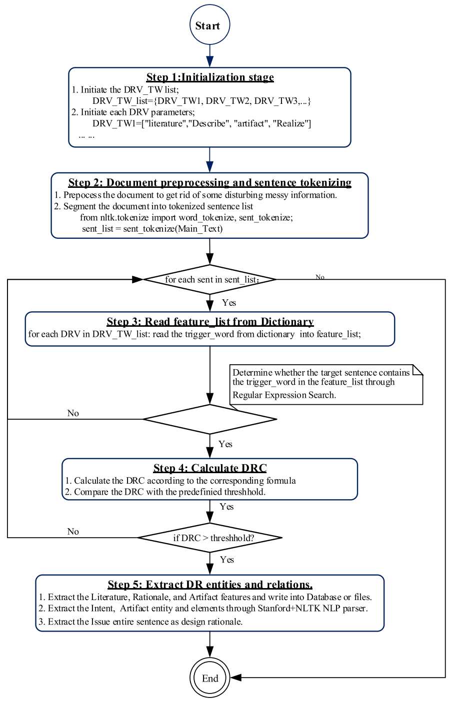


FIGURE 6. Flow chart of DR identification and extraction.

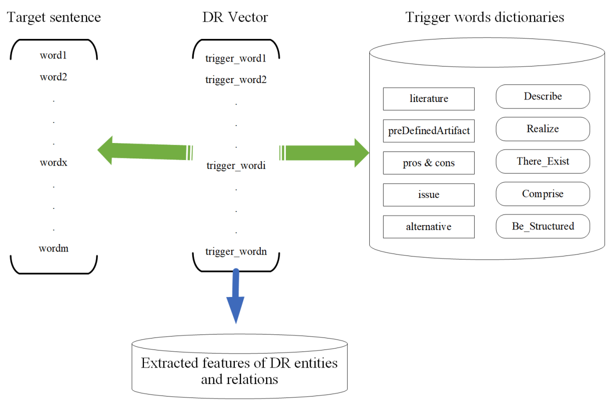


FIGURE 7. DR feature word extraction based on DRVs.

which can be used to establish the citations between patents. Lastly, patent literature usually contains some useful metadata, which can be used to establish the open DR knowledge network between patent documents, such as name of inventor, applicant information, assignee information, and other metadata.

Generally, knowledge fusion focuses on determining whether multiple knowledge acquired from different sources is the same knowledge [68], [75]. For patent literature, the same semantic content is often repeated in a document. The same semantic content is repeatedly extracted, resulting in redundancy of design knowledge. To reduce the redundant knowledge, duplicate DR entities and relationships should be merged, or an "identical_to" relationship should be created between duplicate entities according to specific rules. In this study, linguistic structure of the patent documents is used for knowledge fusion. For example, we assume that there

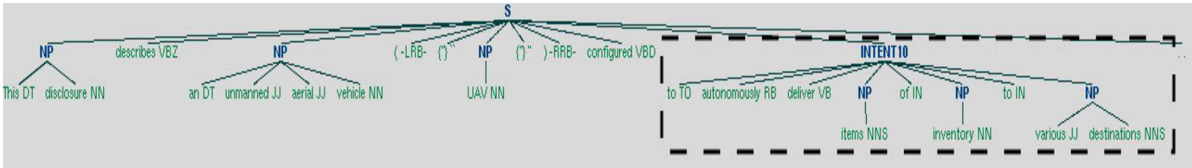
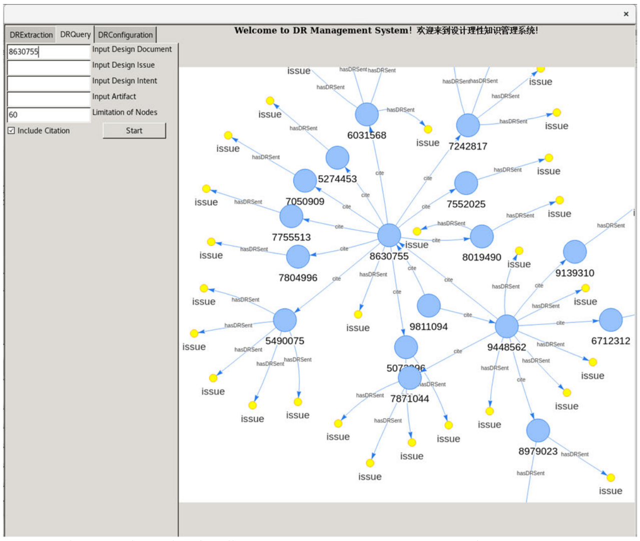


FIGURE 8. Example for Design Intent Extraction.



Note: The numbers in the figure are USPTO patent numbers.

FIGURE 9. Design knowledge network based on the citation of patent documents.

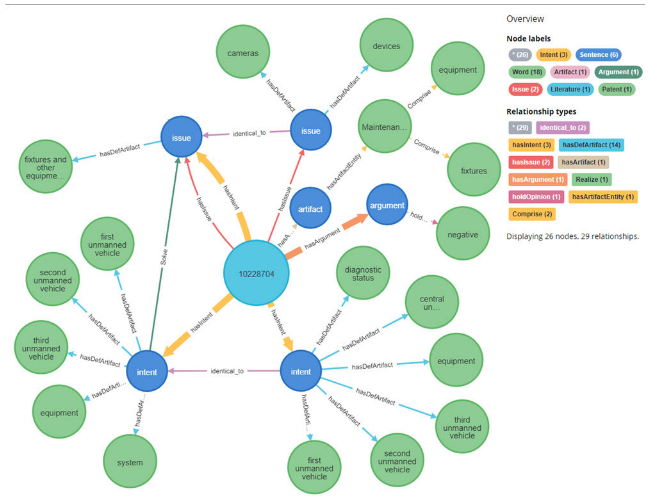


FIGURE 10. Example of knowledge fusion for *issues* and *intents*.

is only one major issue in the ''background'' section of a patent document. Therefore, negative description information can be regarded as a detailed description of the issue in the background section. On the other hand, *issues* occurring in the same BDU or different BDUs of the patent document can be considered to be the same issue. For example, if two *issues* are stated in the same BDU, the two *issues* can be considered the same issue, which can be merged. As shown in Table 6 and Fig. 10.

IV. EXPERIMENTAL STUDY

To verify the feasibility and performance of the proposed method, experiments of patent knowledge extraction have been conducted to achieve automatic DR extraction and automatic design knowledge network construction. Furthermore, two machine learning methods based on BERT and FastText are used in DR sentence extraction experiments, which further verifies the performance of the DRV method in DR sentence extraction under low resource conditions.

A. DATA PREPARATION

Patent selection and data preparation. Using ''unmanned aerial vehicle'' and ''safety'' as search keywords, 1,491 patent documents were collected from the United States Patent Office (USPTO) as sample patents. Considering the large length of the patent text, only part of the content is selected as the sample data, including: patent number, title, abstract, description, background and summary.

Repository preparation. Before extracting DR knowledge, a repository needs to be prepared manually in advance. For the DRV method, the repository includes a DRV-TW dictionary and a DRV pattern library. For FastText and BERT, DR sentences need to be labeled as training data. The open source code annotation tool, Doccano [76], is used to annotate feature words and label sentences, as shown in Fig. 11. Based on B.LIU's Opinion Lexicon [77], some new sentimental n-gram words and corresponding weight indexes are appended to the *argument* TW dictionary, which includes: *positive* words and *negative* words.

The DRV-TW dictionary includes 563 *predefined artifact* words, 39 *alternative* words, 35 *DESCRIBE* words, 57 *issue* words, 26 *literature* words, 4853 *negative* words, 2037 *positive* words, and 52 *realize* words. In addition, nine commonly used DRV patterns are constructed. We prepared 1020 labeled DR sentences as training data. This includes 417 *intent* sentences, 207 *issue* sentences, 122 *alternative* sentences, 113 *positive* sentences, and 328 *negative* sentences. Some sentences may have two more labels. For example, a sentence can be *negative* and *issue* sentences (73), as shown in Table 7. The 1020 labeled data are divided into three groups: training data (60%), validation data (20%), and test data (20%).

The precision and recall of the results depends on the scale of the training data. To evaluate the performance of the three

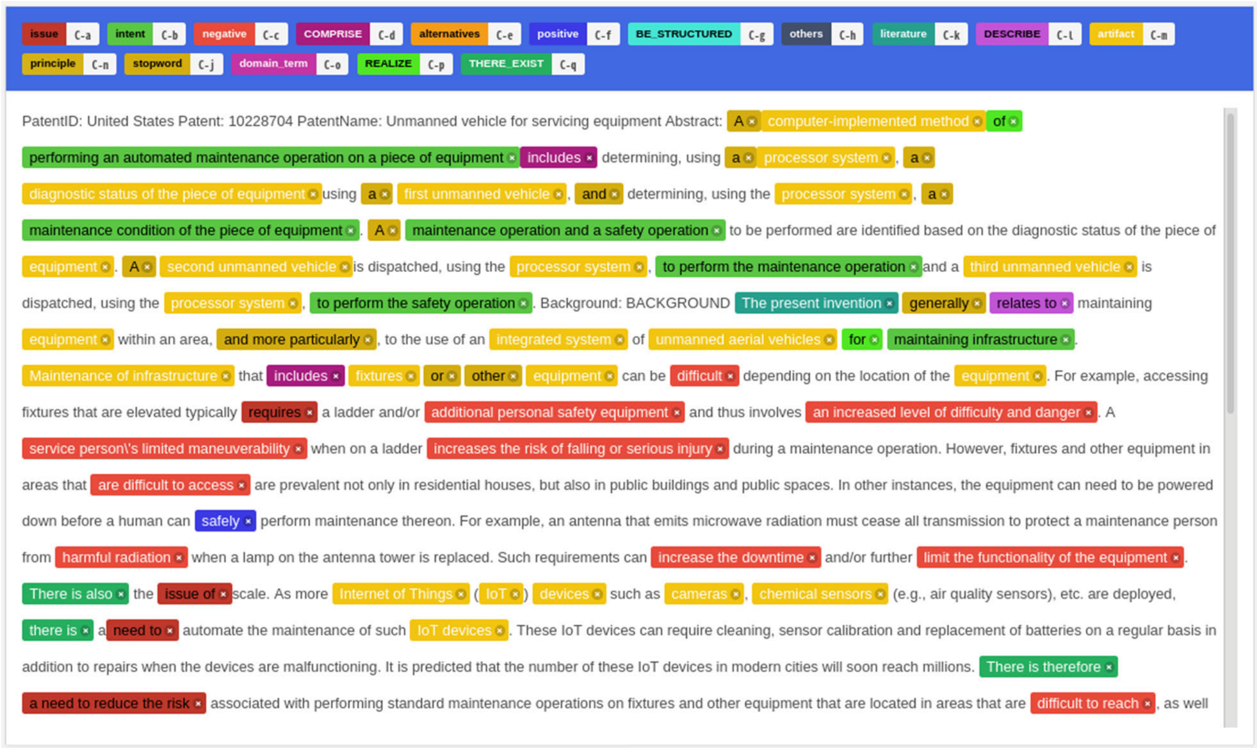


FIGURE 11. **DR annotation of patent documents.**

methods in the case of small corpus, the dictionary data and labeled training data were divided into six groups, accounting for 33%, 40%, 50%, 60%, 80% and 100% of the total corpus, respectively. Using these six sets of data as input, 18 sets of experiments were conducted to evaluate the performance of the three methods under different small corpus.

B. DR KNOWLEDGE EXTRACTION

This study deals with three granularities of DR: sentence-level DR, such as *issues* and *alternatives*; word-level DR, such as *artifacts*, *pros* and *cons*; *Intent* is the DR at the phrase or clause level. For specific levels of DR, appropriate knowledge extraction methods are employed.

1) ARTIFACT

In the experiments, the *artifacts* are extracted by the DRV-TW method and the DRV-DP method, respectively. As a DRV_TW, *predefined artifacts* can be used to not only aid in the identification of DR, but also representing specific DR design objects, which are typically domain knowledge of interest to designers. As a type of DRV-TWs, *predefined artifact TWs* can be extracted based on the DRV-TW method. In addition to *predefined artifacts*, *artifact_components* and *artifact_parts* can be extracted by DRV patterns combined with DP analysis, as shown in Fig. 12 and Fig. 13.

$$NP \ \{<VBG>?<DT>?<JJ.*|VBN>*<NN.*> \\ +<IN>?<DT>?<JJ.*|VBN>*<NN.*>*\}$$

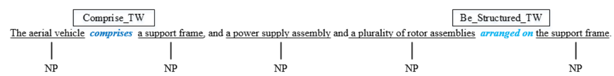


FIGURE 12. **A simple *Comprise* relationship.**

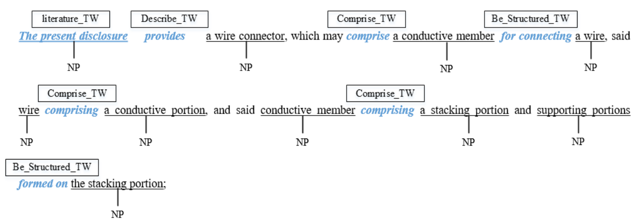


FIGURE 13. ***Literature_Describe_artifact_beStructured_artElement*** **pattern and DP analysis.**

where the noun phrase NP denotes *artifact_components* and *artifact_parts*, which is expressed by the *nltk.grammer* module [78]. Fig. 12 is an example of artifact extraction for a simple DRV pattern (*Comprise* relation pattern) and DP. Fig. 13 shows an example based on the DRV pattern (*literature_Describe_artifact_beStructured_artElement*) and the DP.

2) INTENT

Similarly, the extraction of *intent* is based on DRV pattern and DP. Table 8 shows the context-free grammar of *intent*, which

TABLE 6. **BDU-based knowledge fusion instances for *issues* and *intents*.**

DR type	Instance of DR sentence
Issue 1 stated in "background" BDU	There is therefore a need to reduce the risk associated with performing standard maintenance operations on fixtures and other equipment that are located in areas that are difficult to reach, as well as a need to reduce and in some instances eliminate the dependence on human labor for the maintenance of such devices.
DR type	Instance of DR sentence
Issue 2 stated in "background" BDU	As more Internet of Things (IoT) devices such as cameras, chemical sensors (, air quality sensors), are deployed, there is a need to automate the maintenance of such IoT devices.
Intent 1 stated in "summary" BDU	In one aspect, the present invention provides a system for maintaining equipment within a predetermined area, including a first unmanned vehicle configured to perform a diagnostic evaluation of the equipment, a second unmanned vehicle configured to perform a maintenance operation, and a third unmanned vehicle configured to perform a safety operation.
Intent 2 stated in "summary" BDU	The central control unit is configured to determine a diagnostic status of the equipment in response to data collected by the first unmanned vehicle and dispatch at least one of the second unmanned vehicle to perform a maintenance operation and the third unmanned vehicle to perform a safety operation in response to the diagnostic status of the equipment.

TABLE 7. **Corpus prepared for the three methods.**

DRV-TWs in the dictionary	Labeled DR sentence as training data
predefined artifact TW: 563 *alternative* TW: 39 *Describe* TW: 35 *Issue* TW: 57 *literature* TW: 26 *negative* TW: 4853 *positive* TW: 2037 *Realize* TW: 52	*intent*: 417 *intent*; *alternative*: 17 *intent*; *positive*: 6 *intent*; *negative*: 3 *issue*:207 *issue*; *alternative*: 7 *issue*; *positive*: 30 *issue*; *negative*: 73
	alternative: 122 *alternative*; *positive*: 6 *alternative*; *negative*: 22 *positive*:113 *negative*:328

are defined in the *nltk.grammar* module. Fig. 14 demonstrates the extraction of *intent* based on DRV-DP.

3) PROS AND CONS

The extraction of *pros and cons* adopts a dictionary-based sentiment analysis method. The *argument_TWs*, *pros* or *cons*, contained in a sentence are regarded as feature words representing the argument. The thresholds (T_{pros} and T_{cons}, as shown in Table 4) directly affects the precision and the recall of the results. The larger the T_{pros} setting, or the smaller the T_{cons} setting, the higher precision, but the lower recall.

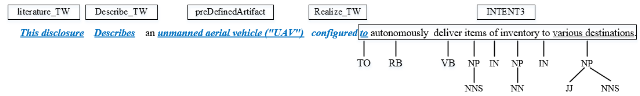


FIGURE 14. **Intent extraction based on DRV and DP.**

TABLE 8. **Grammar representing intent.**

NP	{<DT>?<RB>?<JJ.*|VBN>*<NN.*>+<VBG>?<DT>?<RB>?<JJ.*|VBN|VBG>*<NN.*>*}
INTENT1	{<TO><VB><NP>+<IN>?<NP>?<IN>?<VBG>?<NP>?}
INTENT2	{<TO><VB><CC><VB><NP>}
INTENT3	{<TO><RB><VB><NP><IN>?<NP>?<IN>?<NP>?}
INTENT4	{<IN><VBG><JJ>?<CC>?<VBG|NP>?<NP>+<CC>*<VBG>*<NP>*}
INTENT5	{<IN><NP>?<IN><VBG><JJ>?<CC>?<VBG|NP>?<NP>+<CC>*<VBG>*<NP>*}
INTENT6	{<IN><VBG><IN><JJS><IN>?<NP>*}
INTENT7	{<IN><VBG>+<CC><VBG><NP>*}
INTENT8	{<IN><VBG><RB><RBS><NP><IN><NP><IN>*<NP>*}
INTENT9	{<IN><NP><INTENT1>+}
INTENT10	{<IN><NP><INTENT2>+}

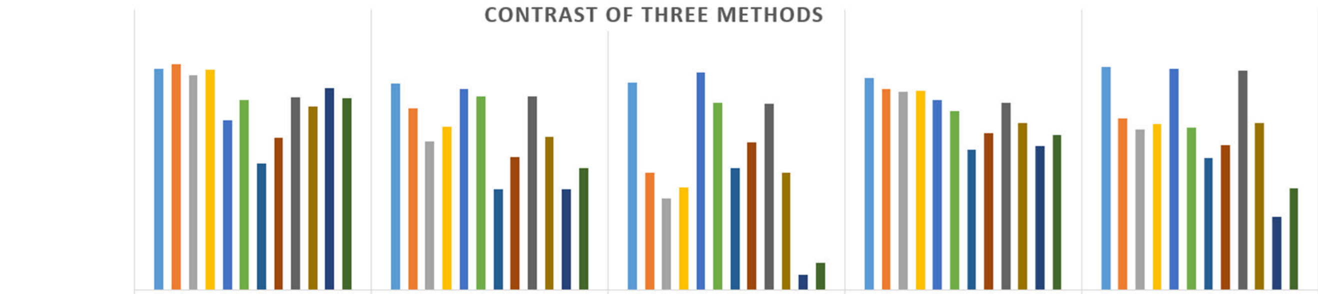
4) ISSUE AND ALTERNATIVE

It is difficult to find the right feature word to represent *issues* or *alternatives*. Hence, the whole *issue* sentence or *alternative* sentence is regarded as an *issue* entity or *alternative* entity. Machine learning methods, such as FastText and BERT, can also easily extract these DRs.

C. RESULT ANALYSIS

Considering that *literature* and *artifacts* are just relevant information for DR, the extraction of *literature* and *artifacts* is not a key objective of the experiment. We focus on the comparative analysis of the extractions of *issues*, *intents*, *alternatives*, *negative arguments*, *positive arguments*, as shown in Fig. 15.

From the analysis of the results, it can be concluded that BERT has low precision and recall when there is a small amount of training data (less than 300 training data), and cannot identify *positive argument*. For *alternatives* extraction, in the case of 40% of the training data, the FastText method could not identify the sentence. This is mainly because the *alternatives* have multiple labels, resulting in low precision in the case of small sample training data. The results show



	INTENT	ISSUE	POSITIVE	NEGATIVE	ALTERNATIVE
BERT_Accuracy	0.946078431	0.882352941	0.887254902	0.906862745	0.955882353
BERT_Precision	0.967032967	0.777777778	0.5	0.859375	0.733333333
BERT_Recall	0.916666667	0.636363636	0.391304348	0.846153846	0.6875
BERT_F-Measure	0.941176471	0.7	0.43902439	0.852713178	0.709677419
DRV_Accuracy	0.725490196	0.857843137	0.931372549	0.81372549	0.946078431
DRV_Precision	0.8125	0.826086957	0.8	0.764705882	0.692307692
DRV_Recall	0.541666667	0.431818182	0.52173913	0.6	0.5625
DRV_F-Measure	0.65	0.567164179	0.631578947	0.672413793	0.620689655
FASTTEXT_Accuracy	0.823529412	0.828431373	0.797356828	0.799019608	0.93627451
FASTTEXT_Precision	0.783018868	0.655172414	0.5	0.714285714	0.714285714
FASTTEXT_Recall	0.864583333	0.431818182	0.065217391	0.615384615	0.3125
FASTTEXT_F-Measure	0.821782178	0.520547945	0.115384615	0.661157025	0.434782609

FIGURE 15. Comparison of DR extraction results.

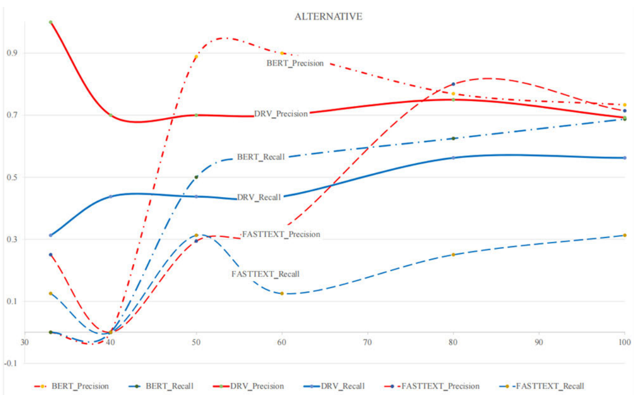


FIGURE 16. Results of *alternative*.

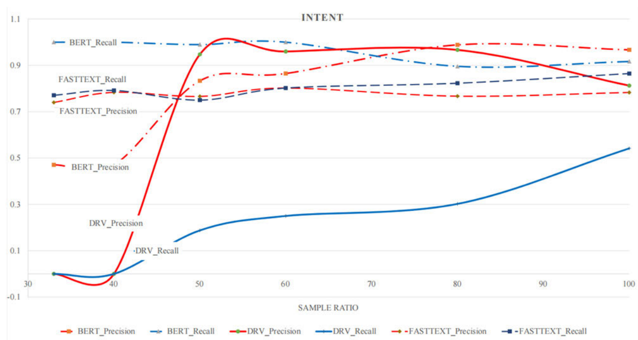


FIGURE 17. Results of *intent*.

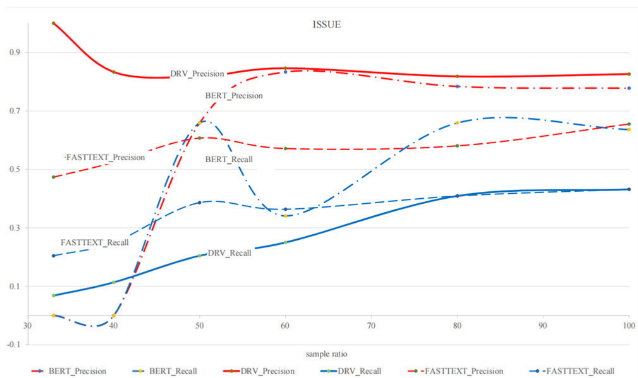


FIGURE 18. Results of *issue*.

that the precision of the DRVS method is relatively high, and the recall will increase significantly as the number of TWs in the dictionary increases. This further verified that the DRV-based DR extraction method is reliable. The recall of the BERT algorithm is relatively high, and the precision will increase as the amount of training data increases. Besides, after the identification of DR, the DRV method can extract some rationale TWs to further construct a KG.

Fig. 16∼Fig. 20. show the results of three methods for DR knowledge extraction in the case of six groups of small corpus.

D. COMPARISON WITH STATE-OF-ART METHOD

To demonstrate the performance of the proposed DRV method, the DRV method is compared with several typical state-of-the-art patent knowledge extraction methods, as shown in Table 9. By comparison, we can conclude that the DRV method has significant advantages in the following aspects:

The DRV method integrates multiple NLP tasks such as knowledge recognition, entity and relationship extraction, and can realize the automatic construction of design knowledge graphs, which is especially suitable for patent knowledge mining. This can further meet the formal representation

TABLE 9. Comparison of patent knowledge extraction for engineering design.

Literature Source	Knowledge Granularity	Extraction method	Knowledge content	Support KG	Support DR
Souili et al. [34]	paragraphs	finite state automata; syntactic pattern	problems, partial solutions, and contradictions	No	No
Fantoni et al. [27]	terms	dependency parser	function, behavior, states/structures of artifacts	Yes	No
Liu et al. [26]	terms	syntactic analysis, WordNet and word vector	effect knowledge	Yes	No
Valverde et al. [4]	terms	Information retrieval; analysis of patents, the identification of pertinent keywords, and the functional decomposition were done manually	problems, functions, physical effects, technology evolution trends and others	No	No
Zuo et al. [32]	terms	Unsupervised method: dependency parser; attention mechanism with BERT	general entities and relations	No	No
Sun et al. [31]	terms	dependency parser	entities, properties and relations of artifacts	yes	No
Liang et al. [30]	paragraphs	term frequencies, language patterns and manifold ranking	issue, solution, artifact	No	Yes
Our approach	terms, phrases, clauses, and sentences	DRV-based method (an improved lexical-syntactic pattern)	artifacts, issues, intents, pros and cons, alternatives, and relations	Yes	Yes

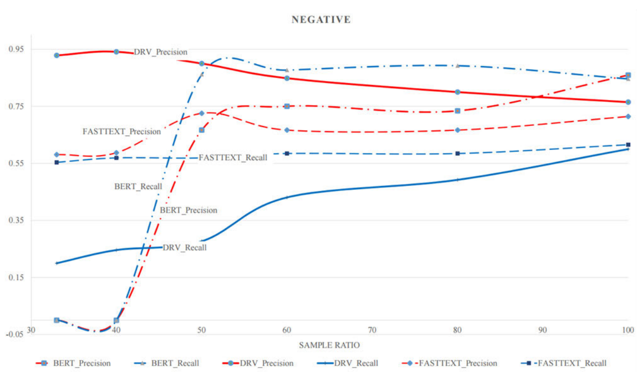


FIGURE 19. Results of ***negative***.

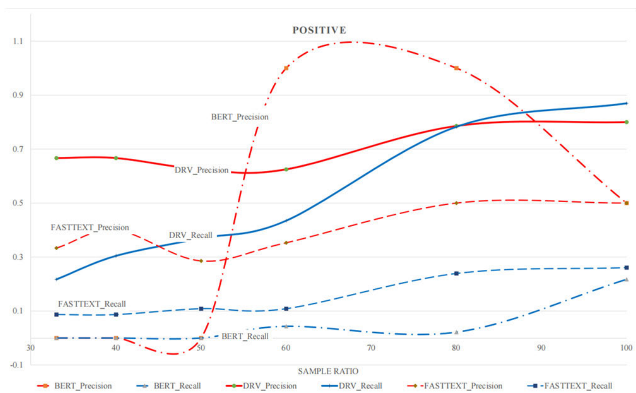


FIGURE 20. Results of ***positive***.

requirements of design knowledge, to realize the designer's rapid retrieval of knowledge.

From a designer's perspective, the DRV approach can provide engineering knowledge of interest to the designer, such as the design *issues*, *intents*, *artifacts*, *pros* and *cons* of the solution. At the same time, the DRV method support the automatic construction of design knowledge graphs, which is more valuable for designers to understand the design context.

V. CONCLUSION

To provide a source of inspiration for innovative design, this study proposes a DRV method to automatically extract DR knowledge, including *literature*, *issues*, *intents*, *artifacts*, *arguments*, *alternatives* and relevant relations, from patent documents. The DRV method includes a DRVS model, which is the theoretical basis of the DRV method. Based on DRVS and lexical-syntactic pattern, we propose concepts such as DRV patterns, DRV-TWs, and *DRC* for the identification of DR sentences. On this basis, a DRV-based DR extraction algorithm is proposed. It integrates knowledge classification, relation extraction, sentiment analysis, named entity recognition, and can be used for DR knowledge identification, DR entity and relation extraction.

One of the contributions of this study is to provide a reliable and convenient method for patent knowledge extraction. The method combines the grammatical features of patent documents to solve several typical NLP tasks. Given the big data nature of patent literature, the precision of DR knowledge extraction task is more important than recall. Experiments show that in the case of low resource corpus, the DRV method has a higher precision for patent design knowledge extraction, which verifies the reliability of the DRV method.

In this study, the DR extracted from the patent contains three levels of granularity: sentence level, phrase level and word level. This simplifies the complexity, while it ignores the different existing forms in which DR knowledge exists. Some DR knowledge exists in the form of words or phrases, some in the form of sentences, and some in the form of multiple sentences. For example, in patent documents, the same question is usually expressed by several negative sentences. Future work on DR knowledge extraction might include research on automatic DR knowledge extraction and KG construction with patent documents of multi-sentences as research objects. Besides, automatic DR knowledge fusion is another issue to be solved.

REFERENCES

[1] R. Bracewell, K. Wallace, M. Moss, and D. Knott, "Capturing design rationale," *Computer-Aided Design*, vol. 41, no. 3, pp. 173–186, Mar. 2009.

[2] R. Setchi and C. Bouchard, "In search of design inspiration: A semantic-based approach," *J. Comput. Inf. Sci. Eng.*, vol. 10, no. 3, Sep. 2010, doi: 10.1115/1.3482061.

[3] L. Liu, Y. Li, Y. Xiong, and D. Cavallucci, "A new function-based patent knowledge retrieval tool for conceptual design of innovative products," *Comput. Ind.*, vol. 115, Feb. 2020, Art. no. 103154, doi: 10.1016/j.compind.2019.103154.

[4] U. Y. Valverde, J.-P. Nadeau, and D. Scaravetti, "A new method for extracting knowledge from patents to inspire designers during the problem-solving phase," *J. Eng. Design*, vol. 28, no. 6, pp. 369–407, Jun. 2017, doi: 10.1080/09544828.2017.1316361.

[5] C. Eckert and M. Stacey, "Sources of inspiration: A language of design," *Design Stud.*, vol. 21, no. 5, pp. 523–538, 2000, doi: 10.1016/S0142-694X(00)00022-3.

[6] G. Yue, J. Liu, and Y. Hou, *Design Rationale Knowledge Management: A Survey*. Berlin, Germany: Springer, 2018, pp. 245–253.

[7] G. Fischer and F. Shipman, "Collaborative design rationale and social creativity in cultures of participation," in *Creativity and Rationale: Enhancing Human Experience by Design*, J. M. Carroll, ed. London, U.K.: Springer, 2013, pp. 423–447.

[8] M. Poorkiany, J. Johansson, and F. Elgh, "Capturing, structuring and accessing design rationale in integrated product design and manufacturing processes," *Adv. Eng. Informat.*, vol. 30, no. 3, pp. 522–536, 2016.

[9] G. Peng, H. Wang, H. Zhang, Y. Zhao, and A. L. Johnson, "A collaborative system for capturing and reusing in-context design knowledge with an integrated representation model," *Adv. Eng. Informat.*, vol. 33, pp. 314–329, Aug. 2017, doi: 10.1016/j.aei.2016.12.007.

[10] F. Mandorli, S. Borgo, and P. Wiejak, "From form features to semantic features in existing MCAD: An ontological approach," *Adv. Eng. Informat.*, vol. 44, Apr. 2020, Art. no. 101088, doi: 10.1016/j.aei.2020.101088.

[11] J. A. Gopsill, H. C. McAlpine, and B. J. Hicks, "A social media framework to support engineering design communication," *Adv. Eng. Informat.*, vol. 27, no. 4, pp. 580–597, Oct. 2013, doi: 10.1016/j.aei.2013.07.002.

[12] G. Arastoopour Irgens, "Connected design rationale: A model for measuring design learning using epistemic network analysis," *Instructional Sci.*, vol. 49, no. 4, pp. 561–587, Aug. 2021, doi: 10.1007/s11251-021-09551-8.

[13] V. Agouridas and P. Simons, "Antecedence and consequence in design rationale systems," *Artif. Intell. Eng. Design, Anal. Manuf.*, vol. 22, no. 4, pp. 375–386, Nov. 2008.

[14] M. Aurisicchio, R. Bracewell, and B. L. Hooey, "Rationale mapping and functional modelling enhanced root cause analysis," *Saf. Sci.*, vol. 85, pp. 241–257, Jun. 2016, doi: 10.1016/j.ssci.2015.12.022.

[15] K. Goucher-Lambert, J. T. Gyory, K. Kotovsky, and J. Cagan, "Adaptive inspirational design stimuli: Using design output to computationally search for stimuli that impact concept generation," *J. Mech. Design*, vol. 142, no. 9, Sep. 2020, doi: 10.1115/1.4046077.

[16] A. J. Trippe, "Patinformatics: Tasks to tools," *World Pat. Inf.*, vol. 25, no. 3, pp. 211–221, 2003, doi: 10.1016/S0172-2190(03)00079-6.

[17] H. Niemann, M. G. Moehrle, and J. Frischkorn, "Use of a new patent text-mining and visualization method for identifying patenting patterns over time: Concept, method and test application," *Technological Forecasting Social Change*, vol. 115, pp. 210–220, Feb. 2017, doi: 10.1016/j.techfore.2016.10.004.

[18] J. Chung, N. Ko, H. Kim, and J. Yoon, "Inventor profile mining approach for prospective human resource scouting," *J. Informetrics*, vol. 15, no. 1, Feb. 2021, Art. no. 101103, doi: 10.1016/j.joi.2020.101103.

[19] M. G. Moehrle, L. Walter, A. Geritz, and S. Müller, "Patent-based inventor profiles as a basis for human resource decisions in research and development," *R D Manage.*, vol. 35, no. 5, pp. 513–524, Nov. 2005, doi: 10.1111/j.1467-9310.2005.00408.x.

[20] S. Kim and B. Yoon, "Patent infringement analysis using a text mining technique based on SAO structure," *Comput. Ind.*, vol. 125, Feb. 2021, Art. no. 103379, doi: 10.1016/j.compind.2020.103379.

[21] S. Altuntas and M. Sezer, "A novel technology intelligence tool based on utility mining," *IEEE Trans. Eng. Manag.*, early access, Aug. 23, 2022, doi: 10.1109/TEM.2021.3101582.

[22] V. Fernandez, "Cross-country concentration and specialization of mining inventions," *Scientometrics*, vol. 126, no. 8, pp. 6715–6759, Aug. 2021, doi: 10.1007/s11192-021-04044-4.

[23] J. Hao, L. Zhao, J. Milisavljevic-Syed, and Z. Ming, "Integrating and navigating engineering design decision-related knowledge using decision knowledge graph," *Adv. Eng. Informat.*, vol. 50, Oct. 2021, Art. no. 101366, doi: 10.1016/j.aei.2021.101366.

[24] G. Ferraro and L. Wanner, "Towards the derivation of verbal content relations from patent claims using deep syntactic structures," *Knowl.-Based Syst.*, vol. 24, no. 8, pp. 1233–1244, Dec. 2011, doi: 10.1016/j.knosys.2011.05.014.

[25] S. Sarica, B. Song, J. Luo, and K. Wood, "Technology knowledge graph for design exploration: Application to designing the future of flying cars," in *Proc. 39th Int. Design Eng. Tech. Conf. Comput. Inf. Eng. Conf.*, Aug. 2019.

[26] H. Liu, W. Li, and Y. Li, "A new computational method for acquiring effect knowledge to support product innovation," *Knowl.-Based Syst.*, vol. 231, Nov. 2021, Art. no. 107410, doi: 10.1016/j.knosys.2021.107410.

[27] G. Fantoni, R. Apreda, F. Dell'Orletta, and M. Monge, "Automatic extraction of function–behaviour–state information from patents," *Adv. Eng. Informat.*, vol. 27, no. 3, pp. 317–334, Aug. 2013, doi: 10.1016/j.aei.2013.04.004.

[28] A. J. C. Trappey, C. V. Trappey, J.-L. Wu, and J. W. C. Wang, "Intelligent compilation of patent summaries using machine learning and natural language processing techniques," *Adv. Eng. Informat.*, vol. 43, Jan. 2020, Art. no. 101027, doi: 10.1016/j.aei.2019.101027.

[29] Y. Liu, Y. Liang, C. K. Kwong, and W. B. Lee, "A new design rationale representation model for rationale mining," *J. Comput. Inf. Sci. Eng.*, vol. 10, no. 3, Sep. 2010, doi: 10.1115/1.3470018.

[30] Y. Liang, Y. Liu, C. K. Kwong, and W. B. Lee, "Learning the 'whys': Discovering design rationale using text mining—An algorithm perspective," *Comput.-Aided Des.*, vol. 44, no. 10, pp. 916–930, 2012.

[31] Y. Sun, W. Liu, G. Cao, Q. Peng, J. Gu, and J. Fu, "Effective design knowledge abstraction from Chinese patents based on a meta-model of the patent design knowledge graph," *Comput. Ind.*, vol. 142, Nov. 2022, Art. no. 103749, doi: 10.1016/j.compind.2022.103749.

[32] H. Zuo, Y. Yin, and P. Childs, "Patent-KG: Patent knowledge graph extraction for engineering design," in *Proc. Design Soc.*, vol. 2, 2022, pp. 821–830, doi: 10.1017/pds.2022.84.

[33] M. M. U. Rathore, M. J. J. Gul, A. Paul, A. A. Khan, R. W. Ahmad, J. J. P. C. Rodrigues, and S. Bakiras, "Multilevel graph-based decision making in big scholarly data: An approach to identify expert reviewer, finding quality impact factor, ranking journals and researchers," *IEEE Trans. Emerg. Topics Comput.*, vol. 9, no. 1, pp. 280–292, Jan. 2021, doi: 10.1109/TETC.2018.2869458.

[34] A. Souili, D. Cavallucci, and F. Rousselot, "Natural language processing (NLP)—A solution for knowledge extraction from patent unstructured data," *Proc. Eng.*, vol. 131, pp. 635–643, 2015, doi: 10.1016/j.proeng.2015.12.457.

[35] G. Guarino, A. Samet, and D. Cavallucci, "PaTRIZ: A framework for mining TRIZ contradictions in patents," *Exp. Syst. Appl.*, vol. 207, Nov. 2022, Art. no. 117942, doi: 10.1016/j.eswa.2022.117942.

[36] M. Lester, M. Guerrero, and J. Burge, "Using evolutionary algorithms to select text features for mining design rationale," *Artif. Intell. Eng. Des., Anal. Manuf.*, vol. 34, no. 2, pp. 132–146, 2020, doi: 10.1017/S0890060420000037.
[37] Z. Kurtanović and W. Maalej, "On user rationale in software engineering," *Requirements Eng.*, vol. 23, no. 3, pp. 357–379, Sep. 2018, doi: 10.1007/s00766-018-0293-2.
[38] B. Liu, *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 2015, p. 383.
[39] N. Eligüzel, C. Çetinkaya, and T. Dereli, "Comparison of different machine learning techniques on location extraction by utilizing geotagged tweets: A case study," *Adv. Eng. Informat.*, vol. 46, Oct. 2020, Art. no. 101151, doi: 10.1016/j.aei.2020.101151.
[40] C.-Y. Chu, K. Park, and G. E. Kremer, "A global supply chain risk management framework: An application of text-mining to identify region-specific supply chain risks," *Adv. Eng. Informat.*, vol. 45, Aug. 2020, Art. no. 101053, doi: 10.1016/j.aei.2020.101053.
[41] J. Kim and C. Lim, "Customer complaints monitoring with customer review data analytics: An integrated method of sentiment and statistical process control analyses," *Adv. Eng. Informat.*, vol. 49, Aug. 2021, Art. no. 101304, doi: 10.1016/j.aei.2021.101304.
[42] A. W. Muzaffar, F. Azam, U. Qamar, and X. Yao, "A relation extraction framework for biomedical text using hybrid feature set," *Comput. Math. Methods Med.*, vol. 2015, Aug. 2015, Art. no. 910423, doi: 10.1155/2015/910423.
[43] S. S. Deepika and T. V. Geetha, "Pattern-based bootstrapping framework for biomedical relation extraction," *Eng. Appl. Artif. Intell.*, vol. 99, Mar. 2021, Art. no. 104130, doi: 10.1016/j.engappai.2020.104130.
[44] Q. Song, Y. Watanabe, and H. Yokota, "Relationship extraction methods based on co-occurrence in web pages and files," in *Proc. 13th Int. Conf. Inf. Integr. Web-Based Appl. Services*. Stroudsburg, PA, USA: Association for Computing Machinery, 2011, pp. 82–89.
[45] H. Yu, H. Li, D. Mao, and Q. Cai, "A relationship extraction method for domain knowledge graph construction," *World Wide Web*, vol. 23, no. 2, pp. 735–753, Mar. 2020, doi: 10.1007/s11280-019-00765-y.
[46] N. Kaushik and N. Chatterjee, "Automatic relationship extraction from agricultural text for ontology construction," *Inf. Process. Agricult.*, vol. 5, no. 1, pp. 60–73, 2018, doi: 10.1016/j.inpa.2017.11.003.
[47] X. Xu and H. Cai, "Ontology and rule-based natural language processing approach for interpreting textual regulations on underground utility infrastructure," *Adv. Eng. Informat.*, vol. 48, Apr. 2021, Art. no. 101288, doi: 10.1016/j.aei.2021.101288.
[48] M. Mintz, S. Bills, R. Snow, and D. Jurafsky, *Distant Supervision for Relation Extraction Without Labeled Data*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2009, pp. 1003–1011.
[49] S. Deng, N. Zhang, H. Chen, C. Tan, F. Huang, C. Xu, and H. Chen, "Low-resource extraction with knowledge-aware pairwise prototype learning," *Knowl.-Based Syst.*, vol. 235, Jan. 2022, Art. no. 107584, doi: 10.1016/j.knosys.2021.107584.
[50] Y. Zhang and Z. Lu, "Exploring semi-supervised variational autoencoders for biomedical relation extraction," *Methods*, vol. 166, pp. 112–119, Aug. 2019, doi: 10.1016/j.ymeth.2019.02.021.
[51] P. Schneider and F. Xhafa, "Machine learning: ML for eHealth systems," in *Anomaly Detection and Complex Event Processing over IoT Data Streams*, P. Schneider and F. Xhafa, eds. New York, NY, USA: Academic, 2022, Ch. 8, pp. 149–191.
[52] A. Mandya, D. Bollegala, F. Coenen, and K. Atkinson, *Frame-Based Semantic Patterns for Relation Extraction*. Singapore: Springer, 2018, pp. 51–62.
[53] M. A. Hearst, "Automatic acquisition of hyponyms from large text corpora," in *Proc. 14th Conf. Comput. Linguistics*, vol. 2. Stroudsburg, PA, USA: Association for Computational Linguistics, 1992, pp. 539–545.
[54] M. Eyal, A. Amrami, H. Taub-Tabib, and Y. Goldberg, "Bootstrapping relation extractors using syntactic search by examples," 2021, [Online]. Available: https://aclanthology.org/2021.eacl-main.128
[55] A. Sun, *A Two-Stage Bootstrapping Algorithm for Relation Extraction*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2009, pp. 76–82.
[56] M. Lazar, D. Militaru, and E. Oancea, "Classifying the lexico-syntactic patterns of semantic relations between two nouns in Romanian language," in *Proc. Int. Conf. Speech Technol. Human-Computer Dialogue (SpeD)*, Oct. 2015, pp. 1–6.
[57] M. G. H. Al Zamil and A. B. Can, "ROLEX-SP: Rules of lexical syntactic patterns for free text categorization," *Knowl.-Based Syst.*, vol. 24, no. 1, pp. 58–65, Feb. 2011, doi: 10.1016/j.knosys.2010.07.005.
[58] G. Zhou, J. Su, J. Zhang, and M. Zhang, *Exploring Various Knowledge in Relation Extraction*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2005, pp. 427–434.
[59] C. Zhang, Y. Zhang, W. Xu, Z. Ma, Y. Leng, and J. Guo, "Mining activation force defined dependency patterns for relation extraction," *Knowl.-Based Syst.*, vol. 86, pp. 278–287, Sep. 2015, doi: 10.1016/j.knosys.2015.06.012.
[60] W. Xu and C. Zhang, "Trigger word mining for relation extraction based on activation force," *Int. J. Commun. Syst.*, vol. 29, no. 14, pp. 2134–2146, Sep. 2016, doi: 10.1002/dac.2897.
[61] R. C. Bunescu and R. J. Mooney, "A shortest path dependency kernel for relation extraction," in *Proc. Conf. Human Lang. Technol. Empirical Methods Natural Lang. Process. (HLT)*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2005, pp. 724–731.
[62] A. Balali, M. Asadpour, R. Campos, and A. Jatowt, "Joint event extraction along shortest dependency paths using graph convolutional networks," *Knowl.-Based Syst.*, vol. 210, Dec. 2020, Art. no. 106492, doi: 10.1016/j.knosys.2020.106492.
[63] S. Xu, S. Sun, Z. Zhang, F. Xu, and J. Liu, "BERT gated multi-window attention network for relation extraction," *Neurocomputing*, vol. 492, Jul. 2021, pp. 516–529, doi: 10.1016/j.neucom.2021.12.044.
[64] S. Yadav, A. Ekbal, S. Saha, A. Kumar, and P. Bhattacharyya, "Feature assisted stacked attentive shortest dependency path based bi-LSTM model for protein–protein interaction," *Knowl.-Based Syst.*, vol. 166, pp. 18–29, Feb. 2019, doi: 10.1016/j.knosys.2018.11.020.
[65] C. J. Fillmore, "Frame semantics," in *Linguistics in the Morning Calm*. Seoul, South Korea: Hanshin Publishing, 1982, pp. 111–137.
[66] H. L. Nguyen, D. T. Vu, and J. J. Jung, "Knowledge graph fusion for smart systems: A survey," *Inf. Fusion*, vol. 61, pp. 56–70, Sep. 2020, doi: 10.1016/j.inffus.2020.03.014.
[67] M. Liu, X. Li, J. Li, Y. Liu, B. Zhou, and J. Bao, "A knowledge graph-based data representation approach for IIoT-enabled cognitive manufacturing," *Adv. Eng. Informat.*, vol. 51, Jan. 2022, Art. no. 101515, doi: 10.1016/j.aei.2021.101515.
[68] B. Zhou, B. Hua, X. Gu, Y. Lu, T. Peng, Y. Zheng, X. Shen, and J. Bao, "An end-to-end tabular information-oriented causality event evolutionary knowledge graph for manufacturing documents," *Adv. Eng. Informat.*, vol. 50, Oct. 2021, Art. no. 101441, doi: 10.1016/j.aei.2021.101441.
[69] M. Ringsquandl, S. Lamparter, R. Lepratti and P. Kröger, *Knowledge Fusion of Manufacturing Operations Data Using Representation Learning*. Berlin, Germany: Springer, 2017, pp. 302–310.
[70] G. Salton, A. Wong, and C. S. Yang, "A vector space model for automatic indexing," *Commun. ACM*, vol. 18, no. 11, pp. 613–620, 1975, doi: 10.1145/361219.361220.
[71] H. A. Simon, *The Sciences of the Artificial*. Cambridge, MA, USA: MIT Press, 1996.
[72] R. G. Thomas and R. Daniel, "Design knowledge and design rationale: A framework for representation, capture, and use," Knowl. Syst. Lab., Dept. Comput. Sci., Stanford Univ., Stanford, CA, USA, Tech. Rep. KSL 90-45, Aug. 1991.
[73] H. Panetto, M. Dassisti, and A. Tursi, "ONTO-PDM: Product-driven ONTOlogy for product data management interoperability within manufacturing process environment," *Adv. Eng. Informat.*, vol. 26, no. 2, pp. 334–348, Apr. 2012, doi: 10.1016/j.aei.2011.12.002.
[74] S. C. J. Lim, Y. Liu, and W. B. Lee, "A methodology for building a semantically annotated multi-faceted ontology for product family modelling," *Adv. Eng. Informat.*, vol. 25, no. 2, pp. 147–161, Apr. 2011, doi: 10.1016/j.aei.2010.07.005.
[75] X. L. Dong, E. Gabrilovich, G. Heitz, and W. Horn, "From data fusion to knowledge fusion," *Proc. VLDB Endowment*, vol. 7, no. 10, pp. 881–892, 2014, doi: 10.14778/2732951.2732962.
[76] H. Nakayama, T. Kubo, J. Kamura, Y. Taniguchi, and X. Liang. *Doccano: Text Annotation Tool for Human*. Accessed: Jul. 16, 2022. [Online]. Available: https://github.com/doccano/doccano
[77] B. Liu, *Opinion Lexicon: A List of English Positive and Negative Opinion Words or Sentiment Words (Around 6800 Words)*. Accessed: Nov. 16, 2022. [Online]. Available: https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon
[78] S. Bird, E. Klein, and E. Loper. *8. Analyzing Sentence Structure*. Accessed: Nov. 16, 2022. [Online]. Available: https://www.nltk.org/book/ch08.html

GAOFENG YUE is currently pursuing the Ph.D. degree with the School of Mechanical Engineering and Automation, Beihang University. He is currently an Associate Professor with the China National Institute of Standardization. He participates in the standardization work of ISO TC184/SC4 industrial data and ISO TC279 innovation management. His research interests include data representation and exchange, design rationale acquisition, and knowledge management.

JIHONG LIU received the Ph.D. degree in mechanical engineering from Tokyo Metropolitan University, Japan, in 1996. He is currently a Professor with the School of Mechanical Engineering and Automation, Beihang University. He has published more than 150 journals and conference papers. His research interests include complex product engineering, knowledge management and knowledge engineering, artificial intelligence in design, and model-based system engineering.

YONGZHU HOU is currently an Engineer with the Beijing Institute of Mechanical and Electrical Engineering. His research interests include digital design and manufacturing, requirements engineering, and knowledge based engineering.

QIANG ZHANG is currently pursuing the Ph.D. degree with the School of Mechanical Engineering and Automation, Beihang University. His research interests include model-based systems engineering and digital thread.