

Research Article

Cite this article: Jing L, Huang M, Li Q, Dou Y, Feng D and Jiang S (2025). Knowledge graph-assisted design-by-analogy: promoting product innovation through structured analogical knowledge retrieval. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, **39**, e20, 1–32
<https://doi.org/10.1017/S0890060425100103>

Received: 30 September 2024
Revised: 30 May 2025
Accepted: 24 June 2025




Keywords:

product conceptual design; knowledge graph; analogical sources retrieval; design-by-analogy; data-driven design

Corresponding author:

Shaofei Jiang;
Email: jsf75@zjut.edu.cn

Knowledge graph-assisted design-by-analogy: promoting product innovation through structured analogical knowledge retrieval

Liting Jing¹ , Mingyang Huang¹ , Qizhi Li², Yubo Dou¹, Di Feng³ and Shaofei Jiang¹ 

¹College of Mechanical Engineering, Zhejiang University of Technology, Hangzhou, China; ²Shining 3D Tech Co, Ltd., Hangzhou, China and ³Industrial Design Institute, Zhejiang University of Technology, Hangzhou, China

Abstract

Design-by-analogy (DbA) is a powerful method for product innovation design, leveraging multidomain design knowledge to generate new ideas. Previous studies have relied heavily on designers' experiences to retrieve analogical knowledge from other domains, lacking a structured method to organize and understand multidomain analogical knowledge. This presents a significant challenge in recommending high-quality analogical sources, which needs to be addressed. To tackle these issues, a knowledge graph-assisted DbA approach via structured analogical knowledge retrieval is proposed. First, an improved function-effect-structure ontology model is constructed to extract functions and effects as potential analogical sources, and six semantic matching rules are established to output entity triplets, and the DbA knowledge graph (DbAKG) is developed. Second, based on the knowledge of semantic relationships in DbAKG, the domain distance and similarity between the design target and the analogical sources are introduced to establish an analogical value model, ensuring the novelty and feasibility of analogical sources. After that, with function as the design target, analogical sources transfer strategy is formed to support innovative solution solving, and TRIZ theory is used to solve design conflicts. Finally, a pipeline inspection robot case study is further employed to verify the proposed approach. Additionally, a knowledge graph-assisted analogical design system has been developed to assist in managing multidomain knowledge and the analogical process, facilitate the adoption of innovative design strategies, and assist companies in providing more competitive products to seize the market.

Introduction

Design-by-analogy (DbA) is a crucial process in the development of human cognition, leveraging knowledge from other domains to explore innovative design inspirations for achieving the design target, thereby stimulating creativity and breaking the inertia of design thinking (Song and Fu, 2022). Through it, designers apply analogical thinking to transfer past design knowledge or cases to the target domain and promote the transfer of knowledge to achieve design goals (Song et al., 2020; Jiang et al., 2022; Li et al., 2023). Conceptual design, as a key stage in new product development, can use DbA to seek mapping relationships between different knowledge domains, thereby stimulating the generation of new design ideas and making it an important approach to design innovation.

Based on prior research on the analogical cognition process (Christensen and Schunn, 2007; Linsey et al., 2008; Casakin et al., 2015; Nie et al., 2022), it primarily comprises three stages: analogical knowledge representation, analogical sources retrieval, and analogical sources transfer. In the analogical knowledge representation, various design knowledge types (e.g., function, effect, structure) are abstractly represented through models like the functional information model (Chen et al., 2020), the concept-knowledge (C-K) model (Liu et al., 2020b), and the function-structure model (Galle, 2009). Based on the models, similarity retrieval of design knowledge is carried out to obtain feasible analogical sources. Existing representation frameworks focus on the standard expression of functions and the transfer of relationships, enhancing analogical efficiency. However, with the widespread use of multidomain knowledge, such as patents and web pages in the design process (Valverde et al., 2017; Liu et al., 2020a), while providing a wealth of inspiration, it also brings challenges in various types of knowledge and forms of expression. Traditional representation framework based on function relationships cannot clearly define the knowledge and its relationships, which limits the feasible improvement of knowledge reuse. To address this, a key research goal of this study is to build an analogical knowledge ontology that represents multidomain knowledge, standardizes knowledge and its relationships in patents and databases, and accelerates the analogical process.

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial licence (<http://creativecommons.org/licenses/by-nc/4.0>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use.

In the analogical source retrieval process, the key to selecting useful analogical knowledge is to accurately evaluate the similarity between analogical sources and design target, which is crucial to increasing the overall value of DbA. Previous studies, such as the imaginary analogy (Hey *et al.*, 2008) and anthropomorphic analogy (Cao *et al.*, 2021), provide a variety of search rules to avoid thinking inertia; however, they do not address how to guide designers in finding reasonable analogical sources. For this reason, text-based retrieval models have been developed to support the extraction of similar analogical sources, such as function similarity (Murphy *et al.*, 2014) and word frequency co-occurrence statistics (Huang and Xie, 2022). Among them, the domain distance between the design target and the analogical sources (Fu *et al.*, 2013) is regarded as an indicator to measure the possibility of generating new ideas and describes the extent of the domain distance between knowledge. Some studies (Srinivasan *et al.*, 2018) point out that remote analogical knowledge can increase designers' creativity and imagination but reduce the feasibility of design results. Furthermore, to enhance the feasibility of analogical knowledge, related studies have defined the semantic similarity of analogical knowledge to emphasize the feasibility of analogical sources (Han *et al.*, 2018a), such as function-behavior-structure (FBS), and the semantic similarity-based model (Guo *et al.*, 2022), and the analogical vector model enhance the adaptability of analogical sources transfer results (Zhu and Iglesias, 2018). Unfortunately, the aforementioned retrieval process neglects the analysis of different knowledge semantic relationships, resulting in subjectivity and ambiguity in the retrieval process. How to comprehensively consider the domain distance and semantic similarity between analogical sources and design targets to filter suitable analogical sources for analogical retrieval, which partially motivates this study. To address this challenge, an analogical value (AV) is proposed and defined as a quantified metric that integrates domain distance and semantic similarity. This metric aims to enhance the precision and objectivity of analogical retrieval processes.

Analogical sources transfer is a mapping process that matches the similar characteristics of analogical sources between the analogical source and the design target, and applies them to DbA (Friesike *et al.*, 2019). Previous studies on analogical source transfer focused on the qualitative representation of knowledge and knowledge processing, but lack in-depth research on the analogical source transfer strategy. Although some knowledge transfer models, such as the retriever analogical reasoning tool (Han *et al.*, 2018b) and the concept combination model (Fu *et al.*, 2015), have been proposed to improve the efficiency of knowledge transfer, their effectiveness is limited by the design experience and the size of the knowledge base. Manually collected fragmented analogical knowledge cannot identify the internal associations between the analogical sources and the design target. When analogical knowledge involves unfamiliar domains, knowledge transfer becomes complex and inefficient. To establish a unified connection between multidomain analogical sources and the design target, analogical characteristics are introduced. These characteristics are defined as core attributes that can be transferred between cross-domain knowledge entities, such as input flows, output flows, scientific laws, and so forth. As the driving force behind analogical transfer, these characteristics are extracted from analogical source knowledge. To organize and utilize the different knowledge levels and its relationships, the knowledge graph technology (Jing *et al.*, 2023) is introduced to systematically organize and represent diverse analogical knowledge and their interrelationships. By utilizing directed graphs,

it delineates the semantic connections between these pieces of knowledge, which is an effective way to bridge the gap in research on the transfer of analogical knowledge across multiple domains.

Unlike traditional approaches that rely on design experience or manual databases to represent and retrieve analogical sources, and employ qualitative reasoning strategies for analogical knowledge transfer, this work introduces a novel innovation design model driven by DbA and supported by the knowledge graph. The aim is to utilize the DbA knowledge graph (DbAKG) to represent and manage multidomain design knowledge and to construct an analogical retrieval process based on the word vector model. Analogical transfer based on knowledge semantic reasoning is then realized to improve the design efficiency and novelty of conceptual schemes (CSs). Finally, the pipeline inspection robot (PIR) design is taken as an example to verify the proposed approach, and two new CSs are provided. Additionally, a knowledge graph-assisted analogical design (KG-AAD) system is developed, which helps designers quickly select design knowledge and improve the efficiency of DbA.

For the sake of clarity, the main contributions of this study include the following three parts:

- (1) By integrating analogical knowledge to construct an improved function-effect-structure (I-FES) ontology model and developing dependency parsing semantic matching rules based on the language technology platform (LTP) (Che *et al.*, 2010), automatic extraction of triplets for DbAKG is achieved.
- (2) Based on DbAKG, an AV model is constructed using the domain distance and similarity between the design target and the analogical sources, leading to the establishment of an analogical sources retrieval model that considers both novelty and feasibility.
- (3) By defining and extracting the analogical characteristics of the function and effect entity, the transfer strategy oriented to function-function (F-F) and function-effect (F-E) are constructed to help designers stimulate the innovative inspiration of multidomain design problems.

The organization of the remaining sections is as follows. Section "Literature review" introduces the correlative research of DbA and the KG technology in product design. Section "Methodology" explains the construction process of DbAKG based on the I-FES ontology and establishes an analogical sources retrieval and transfer strategy to achieve product innovative design. In Section "Case study," the DbA of PIR is taken as an example to verify the proposed approach and compare the novelty and feasibility of different schemes. The study ends with a conclusion and an outlook for future work.

Literature review

This section summarizes relevant literature, including analogical design-driven innovative design and KG-assisted product innovative design, and identifies research gaps in relevant published literature at the end of this section.

Analogical design-driven product innovation design

DbA is an important design approach for product conceptual design, which contains three stages: analogical knowledge representation, analogical sources retrieval, and analogical sources transfer to the design target. By conveying historical design knowledge or solutions, it offers new creativity or solutions to help designers break free from fixed design thinking.

Design knowledge is the basis of promoting product design innovation and the key element of enhancing design thinking innovation, including design principles, technical parameters, function requirements, and case information (Hu et al., 2023). Among them, patents, as the carriers of recording global technical information (Huang and Xie, 2022; Jing et al., 2024), provide a wide source of knowledge for the generation of design schemes. In order to efficiently utilize knowledge, mainstream design ontologies such as FBS and function-effect-structure (FES) (Vermaas and Dorst, 2007; Ma et al., 2016; Chen et al., 2020) have been developed. These ontologies categorize design knowledge into distinct levels, with each level containing only one type of knowledge. For example, function-level covers all function knowledge, defined as describing the actions of a system or component, such as transmitting torque, controlling flow rate, and so forth; effect-level covers all effect knowledge, defined as describing natural phenomena or physical laws, such as Bernoulli's principle, electromagnetic induction, and so forth. This method of dividing knowledge levels through ontology models provides a structured framework for knowledge representation and analogical retrieval. However, when describing the relationship between function and behavior in complex systems, the FBS model is difficult to capture the semantic relationship of design knowledge, which greatly limits the effectiveness of analogy retrieval. In addition, the FBS model neglects the effects and state changes of design objects when associating the aforementioned design knowledge (Wang et al., 2021), which are often the key factors to stimulate innovative thinking. For this reason, He and Hua (2017) introduced the effect knowledge of describing the scientific principles behind the function realization to establish an FES model to improve the adaptability of the design model. The FES model considers a variety of effect principles to inspire designers' innovative thinking and expand the innovative effect of DbA (Song et al., 2017). However, FES is faced with the challenge of representing and utilizing different knowledge levels and their relationships (Beitz et al., 1996), especially in the design problems involving multidomain knowledge (such as biology and electromechanical), and its ability to integrate and apply cross-domain knowledge is limited (Srinivasan and Chakrabarti, 2011; Bhattacharya et al., 2024). Based on this, this study aimed to construct an I-FES knowledge ontology, which introduces the function verb (Fv) and the function noun (Fn), input flow, output flow, and other attributes to represent the relationships between functions and effects in detail, supporting designers to find more usable analogical sources in a wide range of knowledge domains.

Analogy retrieval is a key way to select appropriate analogical data from massive knowledge and reuse historical design cases, including two retrieval modes: search terms-based and questions-based (Jia et al., 2019). In the search terms-based retrieval mode, it can accurately search for the required analogical knowledge, but limits the divergence of analogical thinking. Compared with the former, question-based retrieval mode can expand the divergence of analogical thinking; however, it is difficult to ensure the accuracy of retrieval. To obtain more creative design ideas, retrieving novel and feasible analogical knowledge is a challenge in analogical source retrieval research. To this end, related works (Liu et al., 2022b) adopted knowledge from different fields to stimulate rigid design thinking and generate innovative CSs. Chen et al. (2024) used more than 1600 biological cases in the AskNature¹ online database to provide a source of biological analogies for designers. Luo et al.

(2021) constructed an expert system for innovative design based on knowledge distance and found that the stimulation in the remote domain of data-driven innovative design may generate creative designs. Although the distance between the analogical sources and the target domain is far, more novel design concepts can be provided (Chan et al., 2011); however, the feasibility of the design cannot be guaranteed. To enhance the feasibility of analogical sources, Eilouti (2009) constructed a case-based knowledge management structure to decompose existing design combinations into simple explicit forms, so as to facilitate the classification, analysis, and reuse of their information to obtain new design inspiration. Romero Bejarano et al. (2014) proposed a recursive case-based reasoning model suitable for system design and application in the aviation field. Song and Fu (2019) proposed a method with two retrieval stages: first, retrieving the most similar cases for design reuse, and then, retrieving relevant function units for analogy in the retrieval stage. Liu et al. (2020a) proposed a patent knowledge retrieval method based on function similarity, which can be effectively verified in the design process of ice-breaking equipment. Analogical retrieval considering similarity can obtain cases in a specific field according to specific design problems, most of which come from similar products or similar products, which inhibits the possibility of design innovation. In summary, how to construct an AV that balances the novelty and feasibility of analogy retrieval results is the key to ensure that analogical search results have high quality and strong innovation in design innovation.

Besides, analogical transfer, as an application stage of mapping analogical knowledge to design goals, is a process of solving design problems by using similar knowledge. Liu et al. (2020b) proposed a bionic design model based on C-K theory, normalized the hierarchical knowledge attributes of function-tactics-action structure, and formulated the bioengineering analogical transfer strategy to support the cross-domain mapping of biological prototypes to engineering problems. Li et al. (2023) proposed an analogical knowledge transfer strategy based on structure-mapping function model, which promoted the rapid generation of innovative design schemes. Jia et al. (2019) introduced the relationship-structure-behavior-function ontology model to support the construction of analogical transfer strategies based on semantic similarity, improving the quality of information retrieval. However, the above analogical transfer process depends on the similarity between analogical sources and design target, focusing on the retrieval effects of analogical knowledge (Sarica et al., 2020), and its strategy relies on expert experience while lacking knowledge management rules. To enable cross-level analogical transfer, this study categorizes analogical transfer strategies into two types based on the knowledge levels of the I-FES ontology: same-level F-F analogy (—F-F within the same knowledge level) and cross-level F-E analogy (F-E across different knowledge levels). For example, to address desert water harvesting, the function "collecting water vapor from air" was transferred to the beetle's function "collecting moisture from fog," leading to a biomimetic water collection device. This belongs to analogical transfer within the same knowledge level. To improve water quality, the function "purifying water" was transferred to the "photocatalytic effect," resulting in a nanophotocatalytic water purification system. This belongs to analogical transfer across the different knowledge levels.

For this reason, KG is introduced in this work to represent analogical knowledge and its relations. Knowledge semantic relations are used to support the characteristics extraction and matching between the analogical sources and the design target. Two types

¹<https://asknature.org/>

of transfer strategies are established, namely F-F analogical transfer and F-E analogical transfer, to make analogical source transfer more objective and effective.

Research on KG-assisted product innovative design

KG originated from Google's development of the next generation of intelligent semantic search engine technology (Hao *et al.*, 2021; Zhou *et al.*, 2021), which is essentially a semantic knowledge base based on directed graph. The knowledge storage results are triples, and two kinds of entity-relationship-entity and entity-attribute-attribute value are adopted, where nodes represent entities or concepts, and edges represent their interconnections. Considering the advantages of KGs in representing and reasoning large-scale knowledge in product innovation design, related works (Liu *et al.*, 2022a; Liu *et al.*, 2023b) pointed out that web pages, patents, and other rich design resources can provide important knowledge resources for inspiration. In particular, patents that record massive technical knowledge (Giordano *et al.*, 2023; Jiang *et al.*, 2024), whose unstructured texts contain a lot of design information, can solve the dilemma of poor source of design knowledge.

Simultaneously, natural language processing (NLP) technology provides tools such as word segmentation, syntax analysis, and semantic matching for large-scale patent text mining (Puccetti *et al.*, 2021; Rahman *et al.*, 2021; Tan and Zhang, 2021). Liu *et al.* (2020a) proposed a method of design knowledge extraction based on function base, which used a semisupervised learning algorithm to extract cross-domain function knowledge of patent text. Huang *et al.* (2023) combined patented design knowledge with design contradictions resolution methods to further improve the efficiency and quality of design innovation. It can be seen that the design knowledge in patents is represented in the form of triples (Wu *et al.*, 2021), and knowledge parameters are stored in the KG in the form of attribute values, which is an effective method to promote cross-domain product design innovation by using diversified design. How to establish knowledge ontology accurately and completely is very important for managing different types of design knowledge. For example, Zhang *et al.* (2021) constructed a KG based on the requirement of FBS knowledge ontology to make design knowledge more interpretable and improve knowledge retrieval performance. Wang *et al.* (2021) proposed a context-aware demand stimulation method based on KG to effectively mine potential user demand knowledge. In order to meet the personalized and dynamic needs of users, Zhang *et al.* (2020) built a design knowledge representation framework integrating a case database and KG, which enhanced the scalability and flexibility of the case database and improved the retrieval efficiency of design cases. Haruna *et al.* (2023) combined the KG and patent to solve the problem of tacit knowledge acquisition and matching, and verified it in the case of throttle pedal design. It can be seen that although patents contain rich design knowledge, their text expression is semistructured or even unstructured. How to automatically extract analogical knowledge and construct semantic relations between different knowledge levels is a key problem to solve the difficulty of patent text to support the DbA process.

As tools for associating and visually displaying the relationships between knowledge entities (Shen *et al.*, 2022), KGs can mine potential semantic relationships between knowledge levels to achieve semantic retrieval of knowledge. Zhou *et al.* (2021) developed the service KG and constructed cosine similarity based on Word2vec word vector to calculate text similarity and improved the

accuracy of service knowledge retrieval by using semantic matching of knowledge. Guo *et al.* (2022) integrated machining knowledge into the FBS model and constructed knowledge retrieval rules, and used the fuzzy evaluation model to improve the correlation between retrieval knowledge and design objectives. Zheng *et al.* (2021) built a knowledge ontology model for low-carbon products, integrating case base and KG to support semantic retrieval of design cases, and improved the design efficiency of low-carbon products. Avdeenko and Makarova (2018) established a knowledge representation model integrating case database and domain ontology and used semantic case retrieval to solve the problem of low efficiency and inaccuracy of traditional case database retrieval. Peng *et al.* (2017) conducted a unified representation of explicit and implicit knowledge and designed a knowledge space integrating geometric models, knowledge analyses, and problem-solving strategies to meet the real design requirements in collaborative engineering design. In a word, KGs have technological advantages in designing knowledge representation and inference, not only supporting the visualization of complex relationships between knowledge entities but also providing powerful tool support for semantic analysis and relationship retrieval of innovative design.

According to the summary of the literature, a comparison of the proposed DbA design approach and other design models incorporating knowledge retrieval is presented in Table 1. Based on other observations in the related literature, previous studies have focused on mining analogical knowledge from patents and cases, and evaluating the similarity match between analogical knowledge and design problems typically considers only feasibility or novelty. Some work relies on manual operation or design experience to extract analogical knowledge, resulting in low collection efficiency and low quality. For this purpose, utilizing the KG to retrieve the feasible and novel analogical knowledge is a hot topic to address the innovation of DbA.

Methodology

Figure 1 shows the DbAKG-driven product innovation design framework, which includes the following three parts:

Part 1: I-FES analogical knowledge ontology and its entity triple structure are constructed. Then, an entity recognition method based on dependency syntax analysis and a BiLSTM-CRF training model is developed using patent text, websites, and a historical analogy database as knowledge sources. Six types of knowledge entity relationships are extracted by an NLP tool, and this analogical knowledge is stored and visualized by the Neo4j graph database, ultimately constructing the DbAKG.

Part 2: Functions are determined by analyzing the word frequency and part of speech of technical background words, and the key functions and analogical search terms are obtained using the function clustering model. The shortest path search results obtained from DbAKG describe the domain distance between different search terms. Next, the semantic similarity between F-F and F-E based on the cosine similarity is calculated. On this basis, an AV model is established to screen suitable analogical source knowledge by combining novelty and feasibility.

Part 3: Based on DbAKG, two analogical transfer strategies, F-F and F-E, are constructed to achieve initial solution of analogical design schemes. Based on TRIZ theory, the physical and technical contradictions in the initial DbA schemes are revealed, and the corresponding schemes are provided for different contradictions to obtain the new CSs of DbA.

Table 1. DbA drives the generation of product innovation design schemes

Related works	Data sources	Topics involved in DbA			
		Knowledge management	Feasibility	Novelty	Case study
Fu et al. (2015)	Patents		✓	✓	A device to collect energy from human motion
Srinivasan et al. (2018)	Design project data			✓	Spherical rolling robots
Jia et al. (2019)	Case base	✓	✓		A robot vacuum cleaner
Song and Fu (2019)	Patents		✓		Energy collection device
Luo et al. (2021)	Patents	✓		✓	Water seepage in subway tunnels
Cao et al. (2021)	Patents and websites			✓	Smart natural resource collecting system
Song and Fu (2022)	Patents	✓	✓		An analogical tool (VISION)
Nie et al. (2022)	Patents and case base		✓		Dust removal system for a solar panel
Liu et al. (2023a)	Patents	✓	✓		A pipeline protection device (PipeGuard)
Chen et al. (2024)	Websites	✓		✓	A sustainable product and a portable solar charger.
Proposed approach	Patents, websites, and case base	✓	✓	✓	A worm PIR, a snake PIR

Analogical knowledge ontology model based on I-FES

Construction of the analogical knowledge ontology model

By exploring the relationship of analogical knowledge helps designers to understand and retrieve high-quality analogical sources. For this reason, by referring to the definition of the FES model and incorporating the F-F and F-E analogical relationships, the I-FES model is proposed to improve the retrieval effect of analogical knowledge and provide rich design ideas for DbA. In the I-FES ontology model, the DbA knowledge is defined as three entity dimensions and named as function (F), effect (E), and structure (S), and the analogical relationship between the function and effect levels is added, as shown in Eq. (1)

$$\begin{cases} FES = \{F \cup E \cup S\} \\ F = [F_v] \cup [F_n] \\ E = \langle E_1, E_2, \dots, E_i \rangle \\ S = \langle S_1, S_2, \dots, S_j \rangle \end{cases} \quad (1)$$

where F represents the function, the purpose of product innovation design, and is composed of F_v and F_n , in which F_v and F_n include three levels (as defined in Tables A.1 and A.2), E represents the effect, which is used to reveal how structures, governed by scientific principles, achieve functions by generating observable phenomena, S represents the structure, which constitutes the basic components of a product or a system and acts as the physical carrier for realizing the product's functions.

To organize the entity relationships and attributes in DbAKG, six types of entity relationships are defined to describe the relationships between three entities, as shown in Table 2.

Based on the I-FES ontology model, the standardized relationship between analogical knowledge entities is constructed, and the entity attributes of function, effect, and structure are stored through node attributes. Based on this, the schema of the I-FES ontology-based DbAKG is defined in Figure 2. This schema is designed to represent cross-domain knowledge in engineering design projects in a structured manner. In DbAKG, these relationships are

extracted using automated methods such as entity recognition and relation extraction, ensuring the dynamic mapping of cross-domain knowledge. This enhances the applicability of the method to analogical design by facilitating efficient knowledge retrieval and transfer. Additionally, the use of a unified representation format improves communication among team members and ensures consistency in managing multidisciplinary knowledge.

Design data preparation

Knowledge is sourced from patents, websites, and historical analogy databases, which collectively offer a rich array of multidisciplinary scientific principles and function knowledge. Specifically, functions and structures are derived from the technical background and invention content in patent texts (Jiang et al., 2024); effects are obtained from the scientific effect repository provided by the TRIZ effects webpage² (Chan et al., 2021); and analogical relationships are extracted from the extensive collection of analogical design cases available on the AskNature webpage³ (Chen et al., 2024) and historical analogy databases (Srinivasan et al., 2018; Sarica et al., 2020; Jiang et al., 2022), as illustrated in Figure 3.

Using the construction of the DbAKG for the PIR as an example, three types of data sources were analyzed for building the DbAKG. A total of 985 PIR-related patents published between 2014 and 2024 were downloaded from the patent website [innojoy.com](https://www.innojoy.com)⁴. Subsequently, Python's `os` and `re` modules were used to extract the technical background and invention content from the patents and convert them into text format. The TRIZ effects webpage contains over 1200 scientific effects and offers 175 combinations of F_n and F_v as search terms for retrieving relevant effects. Additionally, the AskNature webpage curates more than 800 biomimetic cases, including over 500 biological functions mapped to their corresponding natural strategies and structures. These resources

²<https://www.triz.co.uk/triz-effects-database>

³<https://www.asknature.org>

⁴<https://www.innojoy.com>.

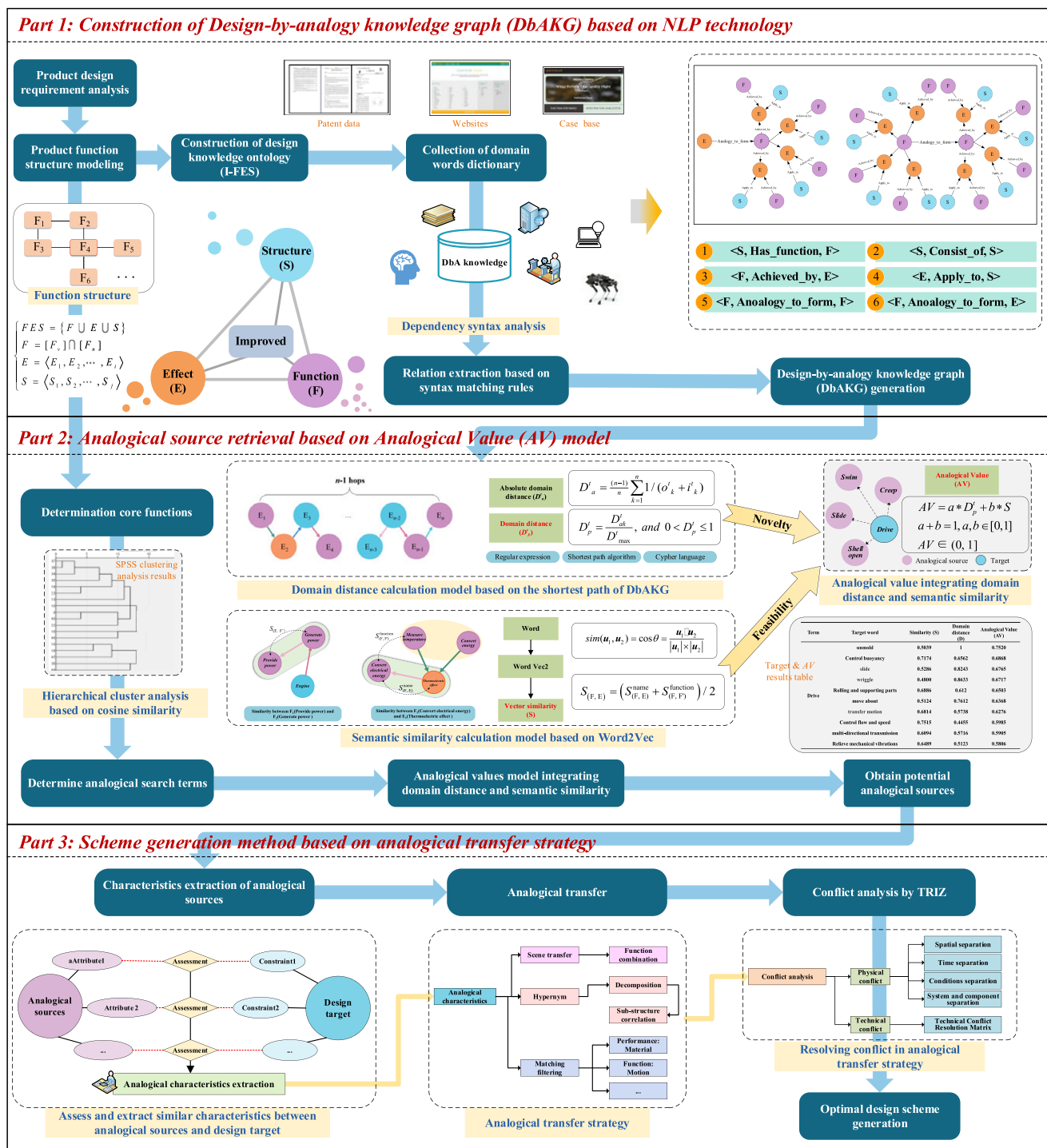


Figure 1. DbAKG-driven product innovation design solution framework.

provide a comprehensive foundation for constructing DbAKG, supporting cross-domain knowledge retrieval and analogical reasoning in engineering design. The detailed statistics and the composition of these three knowledge sources are shown in Table 3.

By extracting function, structure, and effect entities and their relationships from different design data and storing them as <head entities, Relationship, tail entity>, and developing the DbAKG, the specific process is as follows.

Extraction of analogical knowledge entities

Step A1: For function knowledge, the extraction process follows the expression rule of “verb + noun.” First, Fv is derived from a predefined set of abstract verbs in the function basis, as shown in Table A.1. Subsequently, Fn is identified by conducting dependency syntactic analysis of patent texts using LTP, to recognize a noun that has a “VOB” relationship with the Fv. For example, in the sentence “The robotic arm grabs objects,” analysis can extract the

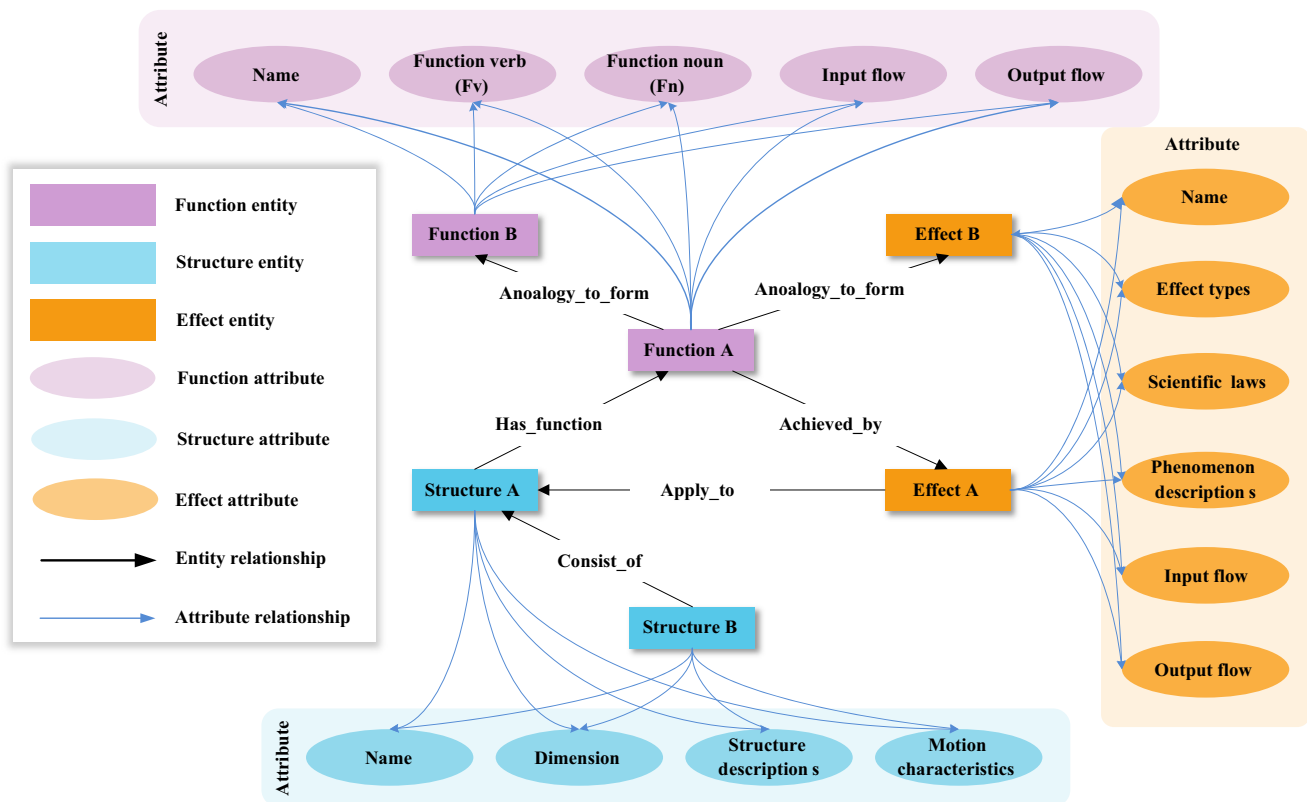


Figure 2. An illustrative schema of the I-FES ontology-based DbAKG.

Table 2. Definition of knowledge entity relationships

No.	Head entity	Definition of dependency relation		Tail entity
		Relationship	Explanation	
1	S	Has_function	This structure has the function	F
2	S	Consist_of	This structure consists of the other structure	S
3	F	Achieved_by	This function is achieved by the effect	E
4	E	Apply_to	This effect applies to the structure	S
5	F	Analogy_to_form	This function is used as an analogy to form the other function	F
6	F	Analogy_to_form	This effect is used as an analogy to form the other function	E

Fv “grab” and the Fn “objects” that satisfy the “VOB” relationship, resulting in the function entity “grab objects.”

Step A2: For effect knowledge, the extraction process focuses on the concrete embodiment of the innovation process in product design, such as biological effect and physical effect. Using the TRIZ effects webpage as the data source, the Octoparse tool⁵ is used to crawl the effect query results, and key attributes such as input and output flows are stored as attributes of effect entities. For example, the effect entity “magnetic adsorption” is extracted, with its effect

type “physical effect,” input flow “magnetic force,” and output flow “adsorption force” registered as entity attributes.

Step A3: For structure knowledge, considering that structure entities are mostly proprietary terms in a specific domain, they have a wide range of domain distribution characteristics. To address this, by referring to the existing work in the subject (Jing et al., 2023), the fine-tuned BERT-BiLSTM-CRF model is used to automatically extract structure entities from patent texts, and extraction models of structure entities in mechanical engineering and electronic engineering are constructed. For example, in the sentence “the drive unit includes a drive motor and a drive wheel,” the structure entities “drive unit,” “drive motor,” and “drive wheel” are extracted.

Extraction of analogical knowledge relations

The DbAKG based on the I-FES ontology defines six types of triples, namely <S, Has_function, F>, <S, Consist_of, S>, <F, Achieved_by, E>, <E, Apply_to, S>, <F, Analogy_to_form, F>, and <F, Analogy_to_form, E>. The extraction process is as follows:

Step B1: For the entity relations of <S, Has_function, F> and <S, Consist_of, S>, the invention content text extracted from patents is preprocessed using an LTP-based dependency syntax analysis tool, including: (a) tokenization, (b) part-of-speech tagging, and (c) dependency relationship (DR) analysis. By constructing five types of semantic matching rules, the entity relationships of “Has_function” and “Consist_of” are extracted, and the predefined triples are obtained, as shown in Table 4. The dependency labels defined by the LTP tool are shown in Table A.3. For example, in the sentence “the sensor module is responsible for monitoring ambient temperature” by applying rule 1, the Fv “monitoring” and the Fn “ambient temperature” are identified as a VOB DR tag, which is defined as a function entity. At this time, using syntactic analysis

⁵<https://www.octoparse.com/>

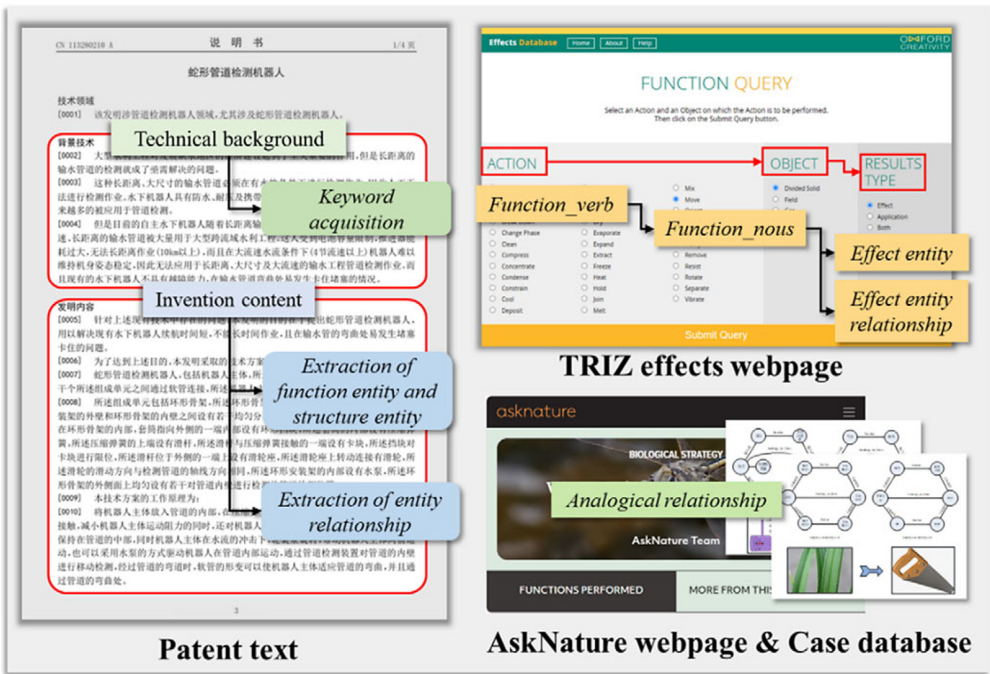


Figure 3. Knowledge sources for DbAKG.

Table 3. Statistics of knowledge data sources for PIR's DbAKG construction

Knowledge types	Data source	Counts
Structure and function	Patent text of PIR	985
Effect	TRIZ effects webpage	1200
Analogical relationship	AskNature webpage and case database	800

according to the law, the SBV tag between the structure entity “sensor module” and the function entity “monitoring” is obtained, which is defined as a “Has_function” relationship between the F and S entities. Then, the triple <sensor module, Has_function, monitoring ambient temperature> is constructed.

Step B2: For the entity relation of <F, Achieved_by, E>, take the TRIZ effect page as the data source to search the effect entity associated with the Fv + Fn. By using the Octoparse tool to construct the triple <F, Achieved_by, E>, capture the F and E

Table 4. Semantic matching rules based on LTP dependency syntax analysis

Extraction rules	Dependency syntactic structure	Triple extraction results
Rule 1		<S, Has_function, F: Fv + Fn>
Rule 2		<S, Has_function, F: Fv + Fn>
Rule 3		<S1, Consist_of, S2> <S1, Consist_of, S3>
Rule 4		<S1, Consist_of, S3> <S2, Consist_of, S3>
Rule 5		<S1, Consist_of, S2> <S1, Consist_of, S3>

Table 5. Example of <F, Achieved_by, E> triple extraction

Function	Relation	Effect	Triple
Absorb divided solid	Achieved_by	Activated alumina	<Absorb divided solid, Achieved_by, Activated alumina>
	Achieved_by	Activated carbon	<Absorb divided solid, Achieved_by, Activated carbon>
	Achieved_by	Adhesive	<Absorb divided solid, Achieved_by, Adhesive>

	Achieved_by	Zeolite	<Absorb divided solid, Achieved_by, Zeolite>

entities that have the entity relationship of “Achieved_by.” Take the function “absorb divided solid” as an example, some search results are shown in Table 5.

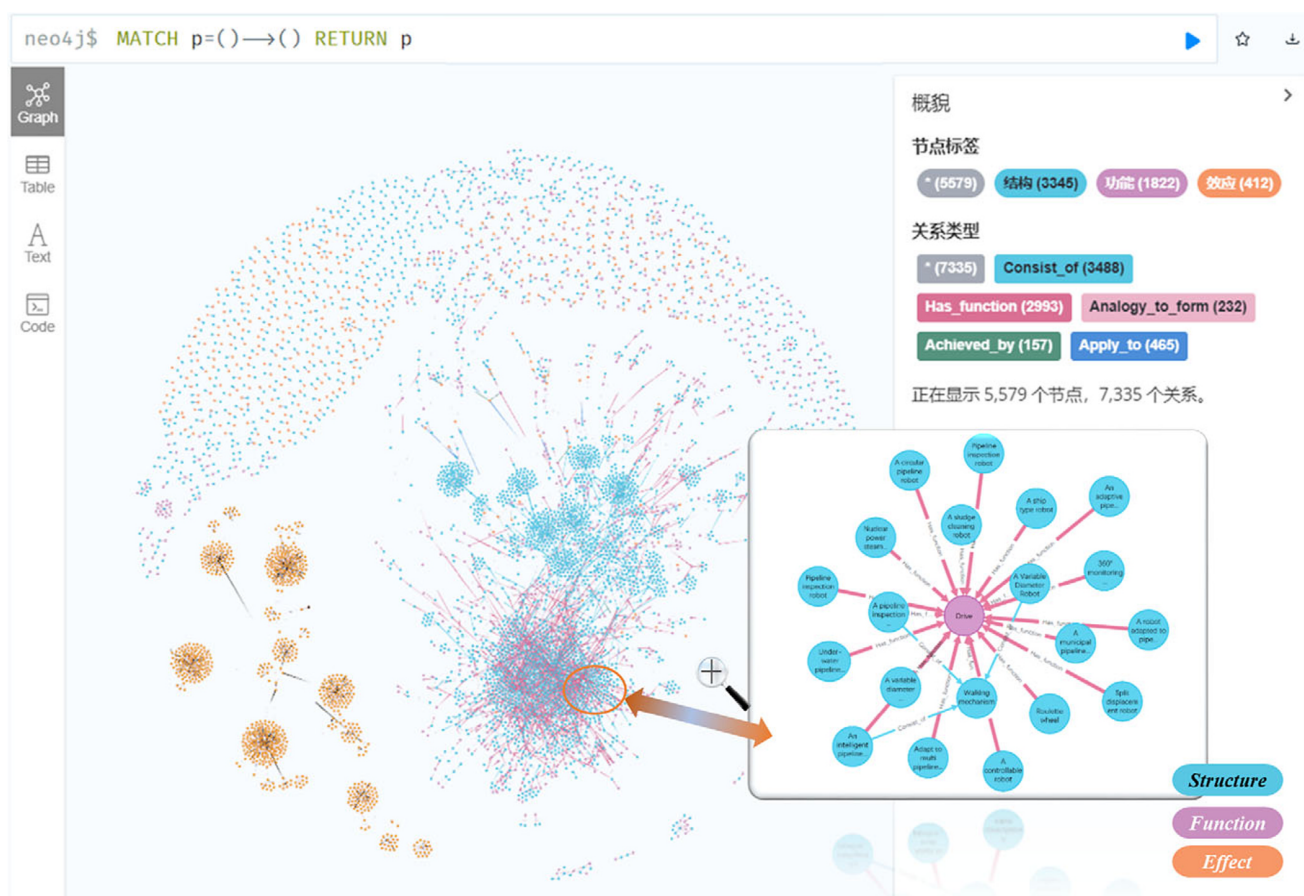
Step B3: For the entity relation of <E, Apply_to, S>, the pre-processed invention content text and effect entities serve as data sources. By using dependency syntax analysis, the subject structure word and its co-occurring effect word in each sentence are extracted, forming the “Apply_to” entity relationship. Then, a triple <E, Apply_to, S> is established based on the retrieved E entity and its co-occurrence S entity. For example, in the sentence “manipulator realizes clamping through friction effect,” the structure entity “manipulator,” the E entity “friction effect” and its relationship

“Apply_to” are extracted, and the triple <manipulator, Apply_to, friction effect> is established.

Step B4: For the entity relations of <F, Analogy_to_form, F>, take the results of function entities extraction as data sources. Identify the second-level classification of the current Fn (Table A.2), replace Fn with new nouns that are in this classification, and use them to establish analogical relationships. For example, the Fv of the “collect exhaust gas” and “collect oil smoke” function are the same, and the Fn belong to the gases in the second-level classification of the Fn. The triple <collect exhaust gas, Analogy_to_form, collect oil smoke> is established. For the entity relations of <F, Analogy_to_form, E>, take AskNature webpage (Chen et al., 2024) and historical analogy database (Srinivasan et al., 2018; Sarica et al., 2020; Jiang et al., 2022) as the source of analogy cases, the triple <F, Analogy_to_form, E> is established through artificial induction, such as triple <attached object, Analogy_to_form, friction effect>.

Generation of analogical KG

To visually represent analogical knowledge entities and their relationships, DbAKG is constructed using Neo4j, a mainstream online graph database. The DbAKG construction process mainly includes two parts: node generation and edge generation. (1) In node generation, each node represents a knowledge entity that can be obtained from the entity recognition model, and the duplicate entity is deleted and the unique node is saved. (2) In edge generation, each edge represents a relationship between entities, all entity edges are created using syntactical matching rules, and DbAKG is generated by combining encoded entity nodes.

**Figure 4.** Entire DbAKG of PIR design showed in Neo4j platform.

Neo4j provides data management and analysis tools and performs search and knowledge reading operations in the semantic web through its high-level query language Cypher, as shown in Figure 4. Different from the traditional KG, the established analogies in DbAKG are formed into three pairs without an aggregation effect, which is conducive to the designer's analogy inspiration and knowledge transfer. For example, through the query function "MATCH p = (n: 'functions' {Name: 'drive'}) <- [r: Has_function] - (m: 'structure') RETURN p" to retrieve the structure entity that satisfies the "drive" function, where the line represents the "Has_function" relationship.

Analogical sources retrieval based on DbAKG

Key function acquisition and search term recommendation

The technical background of the patent covers the technical issues faced by the product development and reflects the purpose or requirement of the product design, that is, the function. Next, the technical background can be used to determine the function search terms and provide the design target for DbA. Among them, Fv can reflect the inventor's description of the product design purpose, and the clustering results of verbs of each category are described as a function requirement, so as to obtain the core function with higher weight as a search term, as described in Figure 5.

First, the relevant patent texts are downloaded from innojoy.com, nonstructural patents are excluded, and the remaining technical background texts are extracted and preprocessed. Dependency syntax analysis is used to identify verbs and their DR tags in sentences, and Fv sets are obtained by locating Fv. Subsequently, the occurrence frequency of each Fv in all texts is defined as word frequency c_i and the weight w_i of corresponding word frequency is calculated by Eq. (2).

$$w_i = c_i / T_k \quad (2)$$

where T_k is the total word frequency of the key verb.

To ensure the satisfaction level of design requirements, a clustering analysis of Fv is conducted to select representative keywords. For this purpose, the verbs are transformed into a low-dimensional word vector space using the Word2Vec⁶ pretrained model, and the set of Fv is clustered using cosine similarity, as shown in Eq. (3).

$$\cos \theta(R_i, R_j) = \frac{R_i \cdot R_j}{|R_i| \times |R_j|} = \frac{\sum_{k=1}^n r_{ik} \times r_{jk}}{\sum_{k=1}^n (r_{ik} - r_{jk})^2} \quad (3)$$

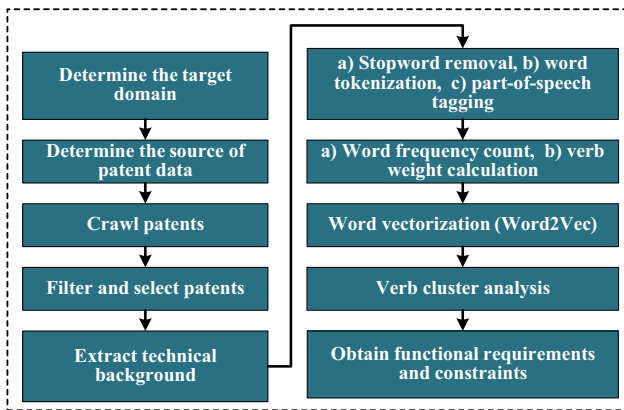


Figure 5. Patent collection and key function acquisition process.

⁶<https://code.google.com/p/word2vec/>

where R_i and R_j represent the n -dimensional vector representations of verbs v_i and v_j , respectively, r_{ik} is the k th component of vector R_i , and r_{jk} is the k th component of vector R_j .

Each function class obtained through cluster analysis represents a group of verbs with a common function purpose. The weight of a function class is the sum of all Fv weights in its set, as shown Eq. (4). The function classes obtained by clustering convey the product design requirements and can be used as terms to search suitable analogical sources.

$$w_i^{cl} = \sum_{k=1}^n w_k \quad (4)$$

where w_i^{cl} is the weight of a function class, and w_k is the weight of the k th function word in the function class w_i^{cl} .

Integrated AV computational model with domain distance and similarity

To retrieve innovative analogical sources, the domain distance is introduced to measure the degree of connection between biological function or effect knowledge and engineering function knowledge. In DbAKG, domain distance is defined as the shortest path between the engineering function (i.e., design target) and the biological function or effect (i.e., analogical source), as shown in Figure 6. The steps for solving domain distance are as follows:

Step C1: Select the knowledge level where the analogical source is located, including the function and effect levels. Use the MATCH function to traverse and retrieve all functions and effects of DbAKG from source nodes.

Step C2: Query multiple shortest paths p between the search term t and the analogical source and obtain the hops $n-1$ and the number of nodes n involved in the shortest path, as shown in Figure 6.

Step C3: According to the nodes obtained in Step C1, query the out-degree o_k^t and in-degree i_k^t of the node k that appears in the shortest path p^t of the search term t , in order to describe the connection relationship between each node and other knowledge, that is, the sparsity of knowledge.

Step C4: According to step C3, obtain the in-degrees and out-degrees of each node queried, calculate the absolute domain distance D_a^t of the shortest path p between the search term t and the node, and use normalization to obtain the domain distance D_p^t between the search term t and all analogical sources, as shown in Eqs. (5 and 6).

$$D_a^t = \frac{(n-1)}{n} \sum_{k=1}^n 1 / (o_k^t + i_k^t), k \in (1, 2, \dots, n) \quad (5)$$

$$D_p^t = \frac{D_{ak}^t}{D_{\max}^t}, \text{ and } 0 < D_p^t \leq 1 \quad (6)$$

where D_{\max} is the $\max\{D_{a1}^t, D_{a2}^t, \dots, D_{an}^t\}$ of the analogy retrieval term to all analogical sources k .

The computational workflow of domain distance is illustrated in Figure 7. By inputting keywords and node types to execute the code, the domain distance results can be derived. The corresponding pseudocode for the domain distance algorithm is also documented in Table A.4.

To enhance the feasibility of analogical sources in specific design problems, cosine similarity based on word vectors is used to calculate the semantic similarity between search terms and analogical sources, providing more suitable analogical knowledge for design target. The specific steps are as follows:

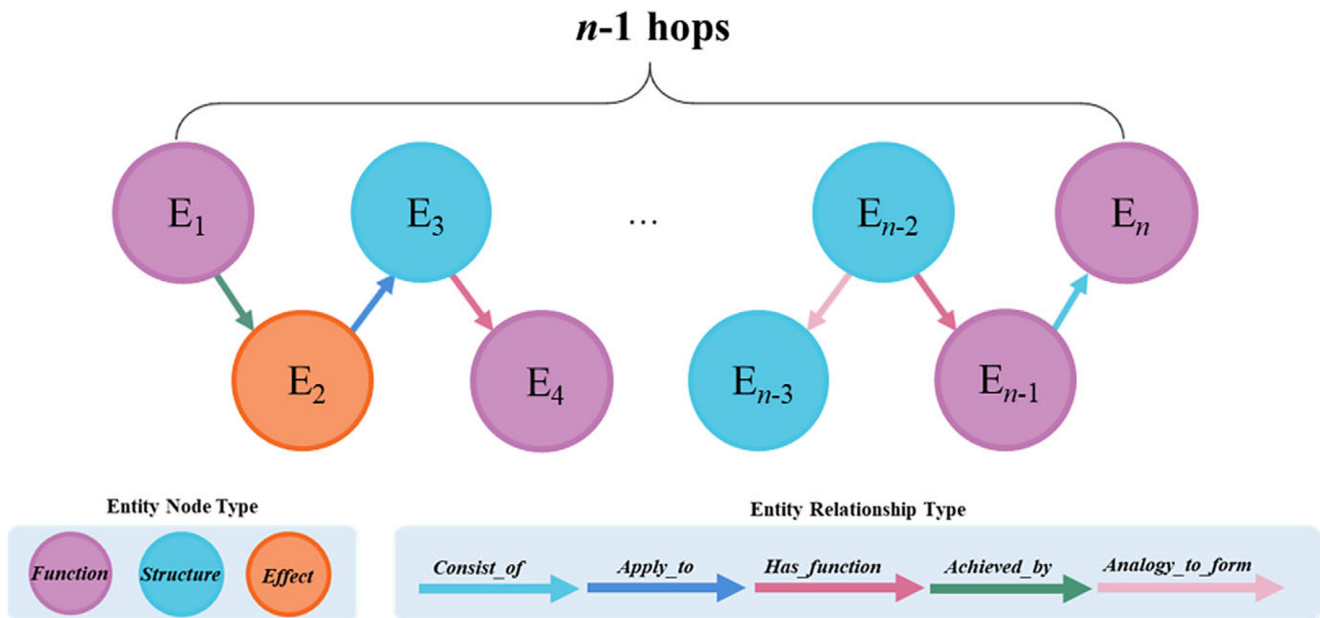


Figure 6. Domain distance calculation model based on the shortest path.

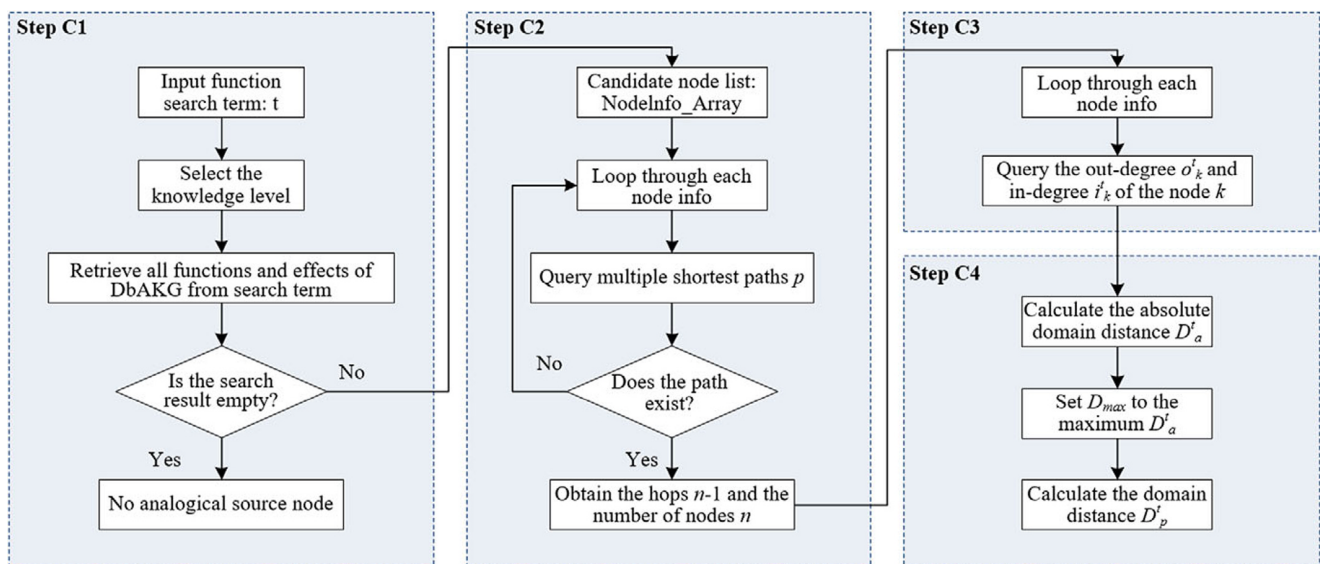


Figure 7. Flowchart for domain distance calculation.

Step D1: Perform preprocessing operations such as tokenization, removing stop-words, and stemming.

Step D2: Use the pretrained Word2Vec word vector model to convert the search term and analogical source text into low-dimensional vectors.

Step D3: For each search term, calculate the cosine similarity between its word vector and the word vectors of all candidate words in the analogical source, as shown in Eq. (7).

$$\text{sim}(\mathbf{u}_1, \mathbf{u}_2) = \cos \theta = \frac{\mathbf{u}_1 \cdot \mathbf{u}_2}{|\mathbf{u}_1| \times |\mathbf{u}_2|} \quad (7)$$

where \mathbf{u}_1 and \mathbf{u}_2 denote word vectors.

In order to calculate the similarity between search terms and various types of analogical sources, this study mainly defines two types of similarity: F-F similarity $S_{(F, F')}$ and F-E similarity $S_{(F, E)}$.

(a) For $S_{(F, F')}$, considering that the search result of search term F and analogical source as a new function F' , and use Eq. (7) to directly calculate the semantic similarity between the two function words.

(b) For $S_{(F, E)}$, considering that the search result of the search term F and the analogical source is a new effect E , it is not only necessary to calculate the semantic similarity between the effect name and the function word, but also to consider the function of the direct association of the effect, so as to fully reflect the actual semantic connection between the two. Thus, $S_{(F, E)}$ includes the similarity $S_{(F, E)}^{\text{name}}$ between effect and function and also includes the similarity $S_{(F, F')}^{\text{function}}$ between the function associated with effect and the search F_n . F' is a function node directly related to E and including “Achieved_by” relationship, as shown in Eq. (8).

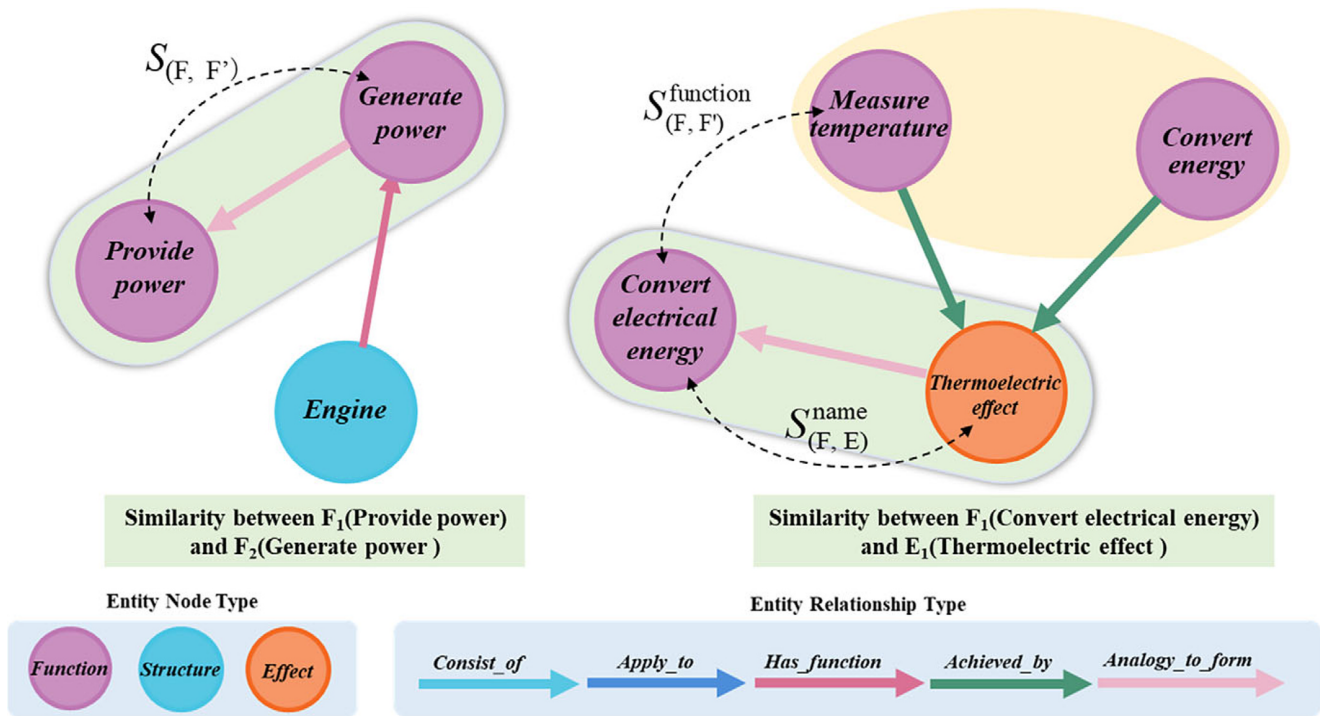


Figure 8. Example of F-F and F-E similarity calculation.

$$S_{(F,E)} = \left(S_{(F,E)}^{\text{name}} + S_{(F,F')}^{\text{function}} \right) / 2 \quad (8)$$

For instance, the similarity calculation between F_1 (convert electrical energy) and E_1 (thermoelectric effect). First, calculate the semantic similarity of F_1 and E_1 , and then calculate the similarity between F_1 and F_2 (measure temperature) which are directly connected to thermoelectric effect and have “Achieved_by” relationship. Eq. (8) is used to calculate the similarity between function and effect, as shown in Figure 8.

Based on the above calculation, an AV model that integrates domain distance and similarity is proposed to quantify the analogy potential of analogical sources toward design targets for the purpose of retrieving analogical sources, as shown in Eq. (9).

$$AV(q_m) = \sqrt{D_p^t(q_m) \times S(q_m)} \quad (9)$$

where $AV(q_m)$ is the AV between the search term q and the m th analogical source knowledge, the larger the $AV(q_m)$, the more likely the analogical source is to be adopted. $D_p^t(q_m)$ is the domain distance, the value range is (0, 1), $S(q_m)$ is the semantic similarity.

Innovative solution seeking based on analogical sources transfer strategy

By extracting and analyzing the analogical characteristics, the analogical transfer strategy is established and the novel design scheme is obtained by solving the analogical source. Then, the TRIZ contradiction resolution method (Jiang and Li, 2016) is used to improve the technical contradiction and enhance the competitiveness of the new CSs.

Extraction of analogical characteristics

Analogical transfer is the matching process between the analogical source and the design target, mapping a new design problem with

the analogical sources. Analogical characteristics represent the core attributes of each analogical knowledge (Zhang et al., 2023), and analogical characteristics (including function and effect characteristics) are extracted from existing analogical sources, as shown in Figure 9. First, compare the F_v and F_n of the function analogical source and the target, confirm whether the phenomenon descriptions of the effect analogical source and the target function are related, and retain the similar items (F_v in Table A.1 of the same level). Second, compare the input and output flows corresponding to the source and target functions, and determine whether they are consistent with the energy flow, material flow, or signal flow.

In addition, the extraction of analogical characteristics follows the principle of priority, systematic, and structure consistency. Specifically, the principle of priority means that when extracting important characteristics of the analogical source and the design target, the analogical characteristics with a high degree of priority similarity are extracted. The systematic principle refers to the consideration of analogical characteristics by the system, not limited to the expression and description of analogical knowledge, but can be associated with factors such as actual scenarios and usage conditions. The principle of structure consistency refers to establishing a one-to-one mapping relationship between the analogical source domain and the target domain.

Analogical sources transfer strategy

Two types of comparative transfer strategies are defined based on the knowledge levels of the analogical source, namely F-F analogical transfer (i.e., transfer within the same knowledge level) and F-E analogical transfer (i.e., transfer across knowledge levels), using the analogy retrieval mechanism to obtain analogical knowledge of functions, effects, and structures.

(a) **Analogical sources transfer strategy for F-F.** Step E1: First, replace the F_v with the first-level F_v in Table A.1 and then refine

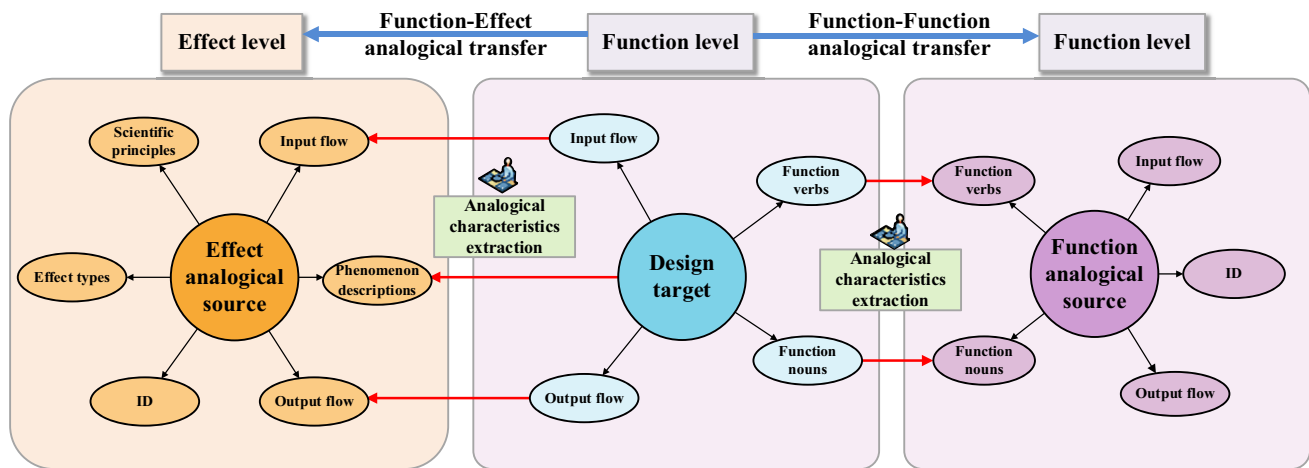


Figure 9. Analogical characteristics extraction strategy.

them into specific third-level Fv. Then, the Fn is replaced with its upper-level word, and the Fn is transferred to the specific Fn of the corresponding level. The relevant transfer information can be queried in Table A.2. For example, taking the ancient water lifting device as an analogical source, its function is described as “conveying water flow,” and the Fn “water flow” belongs to “liquid material.” By transferring it to “solid,” it can be analogized to modern scraper conveyors. This method employs divergent and convergent thinking to abstract the characteristics of the design object. It removes restrictive constraints and introduces specific ones, thereby generalizing the object. This approach facilitates the analogical transfer from the source domain to the target domain.

Step E2: Since the use scene of the product will change with the change of time, space, working conditions and other factors, the function requirements of the product are diversified, and it is necessary to carry out a separate analogical mapping according to the scene. For example, Namibian desert beetles can collect water in arid environments. Drawing on this function, a field water collection device is designed. Through scene analogy, the characteristic is applied to the design task, the connection between the analogical source and the design target is established, and innovative design thinking is stimulated.

Step E3: The relationship between analogical source and design target is not a simple one-to-one mapping, but a complex many-to-one relationship may exist. By combining the functions of analogical sources, significant innovations can be achieved. When combining functions, excessive or insufficient functions should be avoided to ensure that user needs are met. For example, by combining the flight function of an airplane with the driving

function of a car, a flying car can be designed that can both drive on the ground and fly in the air.

(b) Analogical sources transfer strategy for F-E. Step F1: Identify and extract the following key elements characteristic of the effect analogical source: input flows, output flows, key physical parameters, applications, and other distinctive attributes.

Step F2: Abstract the input flows, output flows, and key physical parameters of the effect analogical source into their hypernyms, and map them to the function nouns of the design target, thereby achieving the mapping from effect to function.

Step F3: Identify the application examples and decompose them into distinct substructures. Analyze the specific functions that each substructure performs. Determine how these functions can be adapted to the target domain. Establish a mapping that connects the substructures to their roles in the design.

Some transfers of effect from analogical sources are shown in Table 6. For example, in order to solve the energy consumption problem of the long-distance inspection robot of the submarine pipeline, the analogical source of “solar power effect” is retrieved according to the search term “providing energy.” The input flow (solar energy), output flow (electrical energy), and key physical parameters (power generation) are extracted respectively according to step F1. The search term obtained from step F2 is directly applied to the design goal. Keeping the input flow, output flow, and key physical parameters unchanged, the effect attribute of “solar power effect” is defined as “photovoltaic power generation panel,” and it is mapped to the “power module” design of the submarine pipeline robot.

Table 6. Effect analogical sources transfer example

NO.	Analogical sources	Principle	Analogical characteristics	Design target	Analogical transfer results
1	Venturi effect	Altering flow velocity by changing the cross-sectional area of gas flow	Scientific laws: Bernoulli's principle	Wind turbine	Power generation device enclosure
2	Principle of pepper separation	Pressure difference principle	Scientific laws: Stokes' law	Diamond cutting machine	Pressure differential fracturing device
3	Photoelectric effect	Photoelectric conversion	Input flow: light energy Output flow: electrical energy	Undersea long-distance power module	Photovoltaic power generation module

During the process of analogical transfer, the aforementioned steps are subject to reordering and iteration until an optimal level of innovative stimulation is achieved. Analogical transfer provides designers with many innovative solutions, but there may also be design contradictions that need to be resolved. TRIZ theory provides effective strategies for solving design contradictions by dividing contradictions into physical contradictions and technical contradictions and provides corresponding solutions (Chou, 2021). Physical contradictions occur when mutually exclusive demands are required for the same parameter. Technical contradictions arise when enhancing one aspect of a system inevitably leads to negative impacts on other aspects. Essentially, they represent conflicts between different parameters. The core content of solving physical contradictions is to realize the separation of physical contradictions in the system, including four separation methods: space separation, time separation, condition separation, and system and component separation (Lu et al., 2022). To solve technical contradictions, 40 invention principles can be used to construct a technical contradiction resolution matrix (Wu et al., 2021), and the separation principle can be used to solve physical contradictions to improve the design scheme.

Case study

PIR is a key product for automated maintenance piping systems, including wheel, worm, track, and screw drives. To adapt to complex pipeline systems and flexibly monitor the internal conditions of pipelines, PIR needs to meet the diverse design requirements of bending, reducing diameter, climbing, and so forth. This study takes PIR design as an example, and first, constructs the DbAKG for PIR to acquire multidomain analogical knowledge. Then, the AV model is used to search for the innovative analogical source. Finally, the PIR innovation design is realized by two types of analogical strategies.

Construction of DbAKG for PIR innovative design

Based on the data preparation method in Section “Analogical knowledge ontology model based on I-FES”, a total of 985 PIR patents are retrieved from innojoy.com using the keyword “pipeline inspection robot.” After manual screening and deletion of design patents and software copyrights, 749 patent texts are obtained. The `os` and `re` modules of Python are used to extract the technical background of the patent and the text content of the invention. Then, LTP’s dependency syntax analysis tool is used to preprocess the extracted technical background and invention content text, including: (a) tokenization, (b) part-of-speech tagging, and (c) DR analysis. Finally, the preprocessed text is stored in a txt document to build the dataset of patent text.

Three types of entities are extracted according to steps A1–A3. For the extraction of function entities, the LTP module in Python is used to perform dependency syntax analysis sentences on patent text data, and function entities are extracted from sentences that conform to the rules outlined in step A1. A total of 1822 function entities are obtained. For the extraction of effect entities, according to step A2, the tool Octoparse is utilized to automatically crawl from the database provided by the TRIZ effects webpage, resulting in a total of 412 effect entities being obtained. For the extraction of structure entities, the trained BERT-BiLSTM-CRF model is used to achieve automatic extraction from the invention content text according to step A3. After 314 incorrect structure entities are removed through manual screening, 3345 structure entities are

Table 7. The statistics of the entity nodes in DbAKG

Tag	Has attributes	Counts
<i>Nodes</i>		
Structure	ID, Name, Descriptions	3345
Function	ID, Name, Descriptions, class function	1822
Effect	ID, Name, Descriptions, classification	412
Total		5579

stored in the structure dictionary. The extracted results of the three types of entities are stored in the Neo4j graph database and node information is established. The node statistics of PIR design in DbAKG are shown in Table 7.

Six entity relationships are extracted according to steps B1–B4. On the basis of step B1, syntactic matching rules (rules in Table 4) are used to extract triples satisfying “Consist_of” and “Has_function” from the text of the invention content. Taking rules 1 and 3 as an example, in the sentence “robot cleaning pipe,” there is an “SBV” relationship between “cleaning” and “robot,” and a “VOB” relationship between “cleaning” and “pipeline.” Create triples <Robot, Has_function, cleaning pipeline>. In the sentence “The manipulator consists of a driver and a gripper,” “manipulator” is the subject of the sentence, and there is a “COO” relationship between “driver” and “gripper.” According to rule 3, create the triples <Manipulator, Consist_of, Gripper> and < Manipulator, Consist_of, Driver>, as shown in Figure 10.

Six types of triples are extracted by extraction, and then manual screening is carried out to eliminate repeated triples or triples with obvious extraction errors. A total of 7335 triples are matched, and the entity relationship statistics are shown in Table 8.

Figure 11 shows a DbAKG example of PRI design stored on the Neo4j platform. It can be found that the function node is often the starting node of each knowledge relationship link, and the effect node is the medium connecting each function and structure. Therefore, the analogical transfer is based on function. This approach establishes two analogical processes: F-F transfer and F-E transfer. It also explores the feasibility and potential application scenarios of these analogical approaches involving function and effect. It provides an analysis basis for the potential analogy possibility and application value of DbA.

Retrieval of analogical sources for PIR innovation design

Based on the analogical source search term acquisition method in Section “Analogical sources retrieval based on DbAKG,” Fv of each patent technology background text are extracted, their word frequency and relative weight are calculated, and the top 50 technical keywords with weight are obtained, as shown in Table A.5. Using word2vec training (parameter set: sg = 1, word, size = 192, windows = 8, min_count = 5) calculates cosine similarity of the top 50 technical keywords and cluster analysis, as shown in Figure A.1. Then, verbs are divided into F₁ (inspect), F₂ (drive), F₃ (clean stains), F₄ (remote control), F₅ (easy maintenance), F₆ (adapt to diameter), and F₇ (turn), and Eq. (4) is used to calculate the relative weights of various Fvs, as shown in Table 9. Considering that the realization of F₁ and F₇ depends on the function design effect of F₂, F₃ and F₄ are not necessary functions for pipeline inspection tasks, and the function solving effect depends on the innovation of information technology or optimization algorithms. F₅ is more concerned with its long-term operation efficiency and cost and has no direct impact on

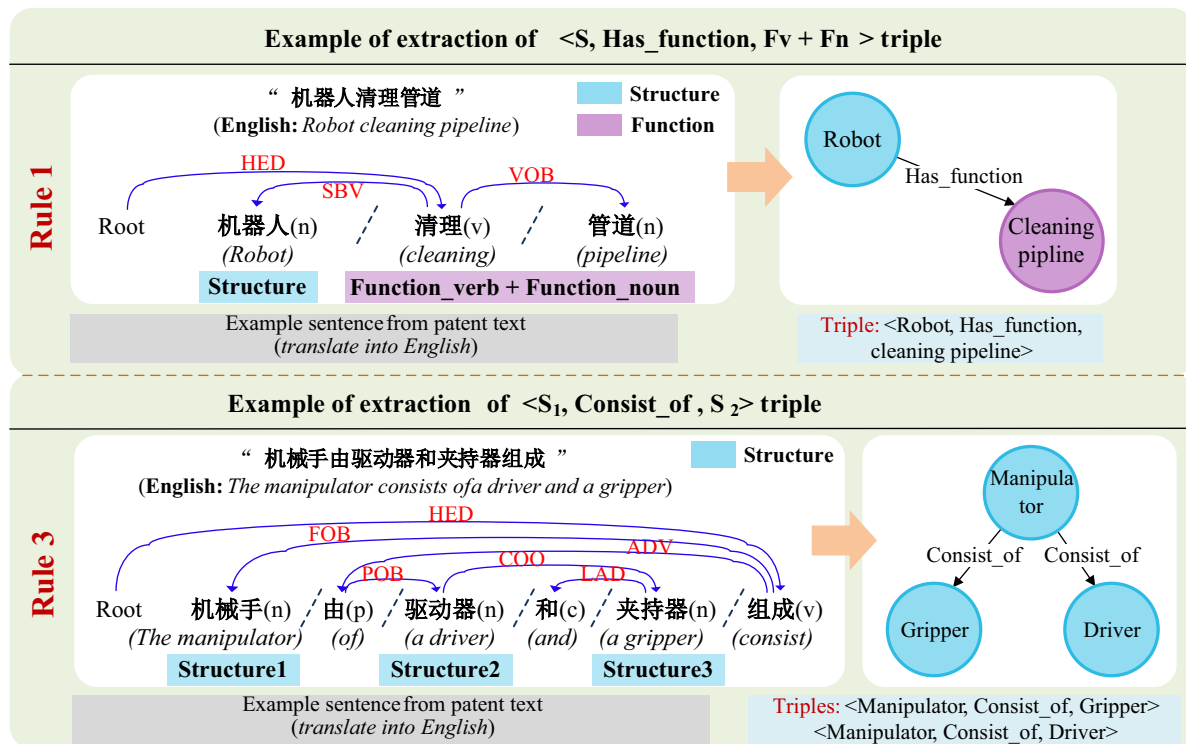


Figure 10. Examples of triples extracted based on rules 1 and 3.

Table 8. The statistics of the entity relations in DbAKG

Tag	Has property	Counts
<i>Relations</i>		
Consist_of	<Structure, Consist_of, Structure>	3488
Has_function	<Structure, Has_function, Function>	2993
Apply_to	<Effect, Apply_to, Structure>	465
Achieved_by	<Function, Achieved_by, Effect>	157
Analogy_to_form	<Function, Analogy_to_form, Function>	128
Analogy_to_form	<Function, Analogy_to_form, Effect>	104
Total		7335

the design scheme innovation of pipeline inspection. Next, F_2 and F_6 are used as the search terms for analogical transfer to complete the CS design of analogical transfer and improve the design effect of PIR.

Analogical sources retrieval based on F-F

According to steps C1–C33 described in Section “Analogical sources retrieval based on DbAKG,” F_2 is selected as the search term for F-F analogical transfer. First, obtain the IDs of all verb nodes from DbAKG regarding PIR. Next, use the MATCH function (e.g., MATCH p = shortestPath (n1: ‘function’{name: “move”})*-(n2: ‘function’{name: “rive”} RETURN p) to search the shortest path p from all function nodes in the graph to F_2 node. Finally, record the Node_n ID of each node in the shortest path p , as shown in Table 10.

For example, the shortest path search process of “drive-open shell” uses the domain distance model to obtain a path with 13 nodes and 12 relationships, with hops of 12, as shown in Table 10. The path p is “drive (ID: 2274), walking mechanism (ID: 4277), walk (ID: 2275), mechanical crab (ID: 1060), crab (ID: 2179), crab robot (ID: 1013), worm (ID: 3362), mechanical starfish (ID: 2135), starfish (ID: 1041), sucker robot (ID: 975), clam (ID: 2151), clam opener (ID: 1055), open shell (ID: 3045),” which includes the “Has_function” and “Analogy_to_form” relationship types, as shown in Figure 12.

Then, the sum of the out-degree and in-degree of each node is calculated, respectively, which are labeled as follows: “drive (54), walking mechanism (28), walk (9), mechanical crab (2), crab (3), crab robot (2), worm (7), starfish (2), mechanical starfish (3), sucker robot (4), clam (2), clam opener (2), shell open (1).” Based on Eqs. (5 and 6) in step C4, the absolute domain distance of “open shell” is calculated as $D_a^{drive} = 12/13 \times (1/54 + 1/28 + 1/9 + \dots + 1/2 + 1) = 4.3614$. Then, the domain distance D is calculated as $D_p^{drive} = 4.3614/D_{max} = 1$. Similarly, the domain distances of the other function analogical sources are calculated. Due to the limitation of the article length, only the top 10 function analogical source domain distances are shown in Table 11.

According to steps D1–D3, the semantic similarity $S_{(F, F')}$ between function words and search terms in Table 10 is calculated using Eq. (7), and the AV is calculated using Eq. (9). For example, the semantic similarity between “open shell” and “drive” is calculated as $S_{(open\ shell, drive)} = sim(u_{open\ shell}, u_{drive}) = 0.5039$. According to Eq. (9), the AV between “open shell” and “drive” is calculated as $AV(open\ shell, drive) = \sqrt{1 \times 0.5039} = 0.7099$. The results of the top 10 analogical sources with AV are shown in Table 12.

The top 10 function analogical sources for AV are “shell open,” “control buoyancy,” “slide,” “rolling support parts,” “worm,” “swim,” “directional transfer,” “regulate flow and speed,” “cut vibration” and

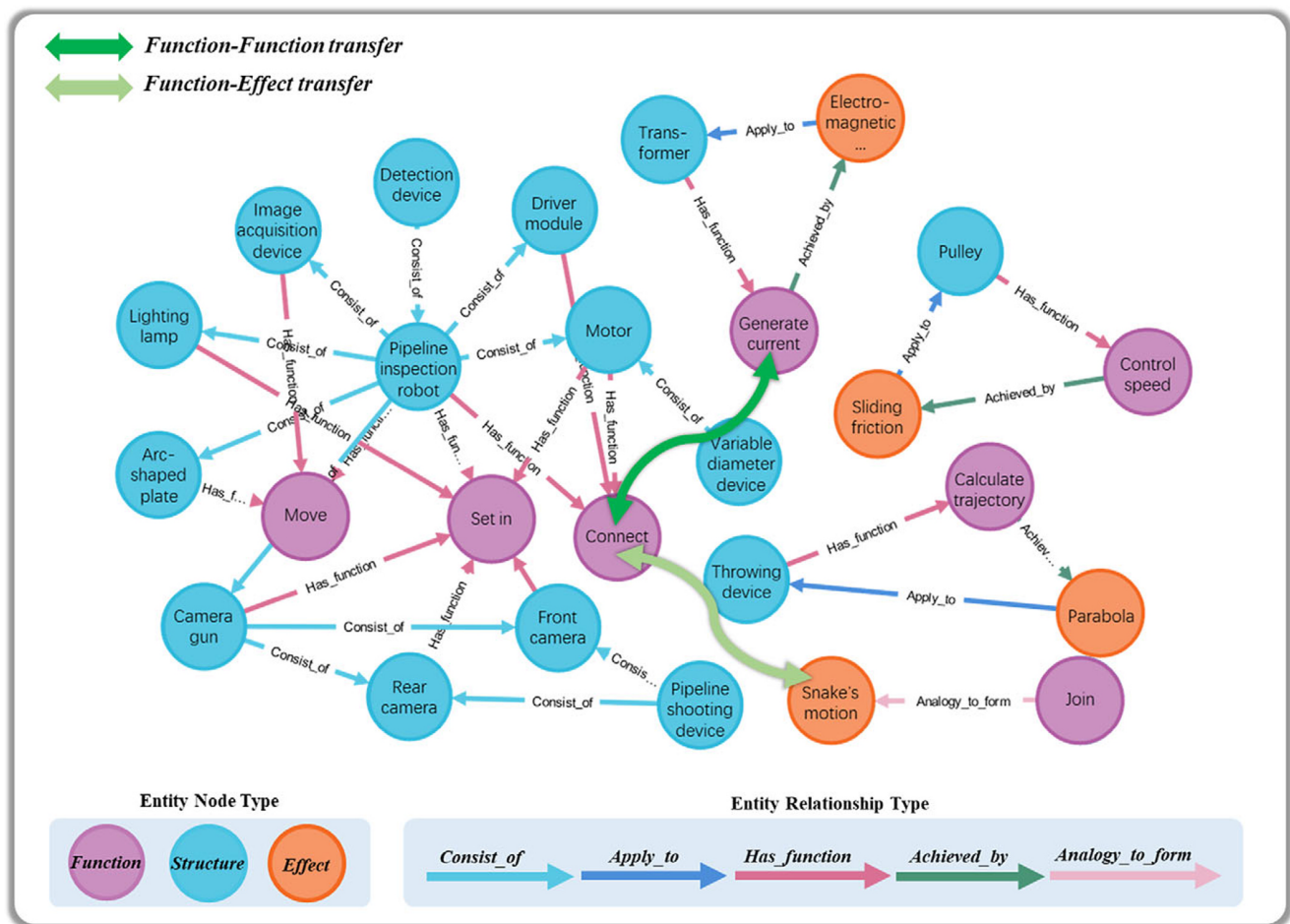


Figure 11. An illustration of DbAKG stored in the Neo4j platform (translated from Chinese into English).

Table 9. Function clustering results

Function class	Function name	Weight (w_{ij}^c)	Fv
F ₁	Inspect	0.2729	Inspect, collect, transmit, identify, measure, detect, monitor, and examine
F ₂	Drive	0.2568	Crawl, transport, convey, tow, run, walk, drive, and worm
F ₃	Clean stains	0.2237	Clean, dredge, drain, leak, block, obstruct, break, damage, destroy, impair, sustain damage, explode, and corrode
F ₄	Remote control	0.0967	Remote control, carry, support, and control
F ₅	Easy maintenance	0.0924	Replace, maintain, repair, fix, restore, push, and weld
F ₆	Adapt to diameter	0.0404	Overcome, obstacles, extend, and adjust
F ₇	Turn	0.0171	Turn, slide, and bend

“connect.” By considering the relevance of the analogical sources to the PIR design, five function analogical sources (marked in bold in Table 12) are selected: “control buoyancy,” “slide,” “worm,” “rolling support parts,” and “swim.” The attribute information of these five function nodes is shown in Figure 13.

Analogical sources retrieval based on F-E

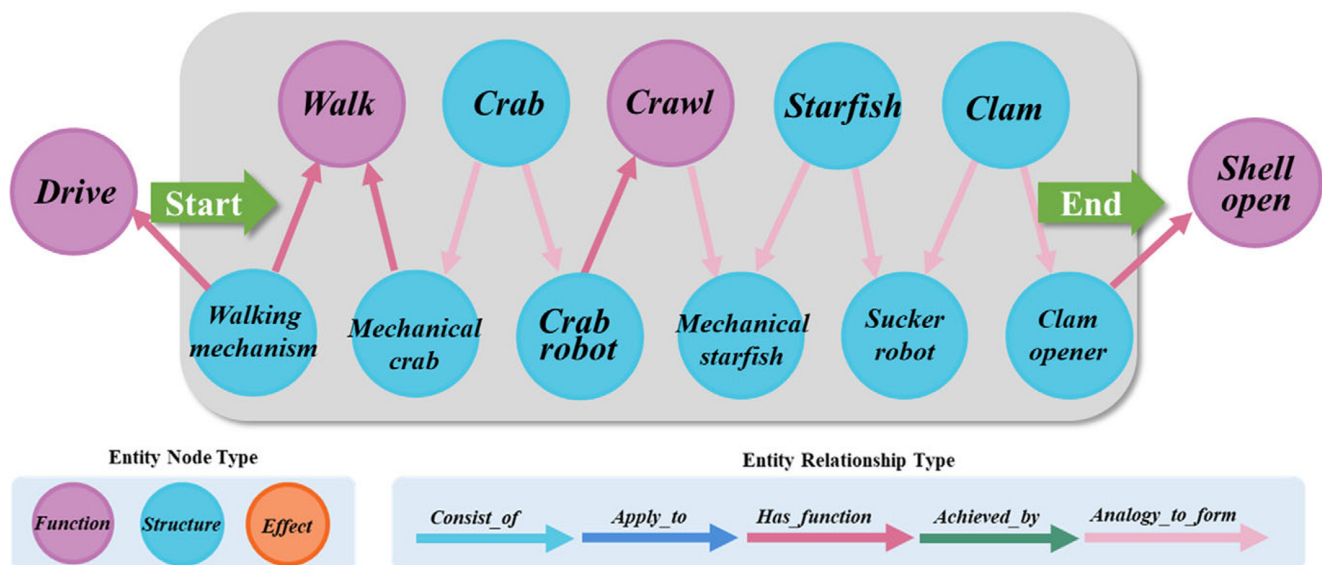
According to steps C1–C3 described in Section “Analogical sources retrieval based on DbAKG,” F₆ is selected as the search term for F-E analogical transfer. First, obtain the IDs of all verb nodes from DbAKG regarding PIR. Next, use the MATCH function to search the shortest path p from all function nodes in the graph to the

“adaptive diameter” function node. Finally, record the Node_n ID of each node in the shortest path p , as shown in Table 13.

For example, the shortest path search process of “adapt to diameter-Archimedes principle” uses the domain distance model to obtain a path with 13 nodes and 12 relationships, with hops of 12, as shown in Table 13. The path p is “adapt to diameter (ID: 2328), drainage pipeline robot (ID: 80), connect (ID: 2256), auxiliary shaft (ID: 4200), bite (ID: 2280), pliers (ID: 982), crocodile (ID: 2143), mechanical crocodile (ID: 1030), swim (ID: 3389), shark (ID: 1052), mechanical shark (ID: 2152), clam opener (ID: 632), open shell (ID: 3493),” which includes the “Apply_to,” “Has_function,” and “Analogy_to_form” relationship types, as shown in Figure 14.

Table 10. Search results for the shortest path nodes of F_2 in DbAKG

Search term	Function analogical sources	Node count	Node ID of the shortest path			
			Node ₁	Node ₂	...	Node _n
Drive	Open shell	13	2274	4277	...	3405
	Worm	11	2274	4277	...	3411
	Slide	12	2274	4355	...	3361
	Swim	12	2274	4355	...	3336
	Control buoyancy	12	2274	4355	...	3203
	Rolling support parts	9	2274	175	...	3156
	Fly	8	2274	4354	...	3383
	Transfer motion	11	2274	175	...	3121
	Directional transfer	9	2274	5024	...	3109
	Cut vibration	7	2274	5024	...	3107
	Control motion	11	2274	175	...	3088
	Hydraulic rotary joint	7	2274	1755	...	2979
	Mechanical link	9	2274	24	...	3123
	Regulate flow and speed	8	2274	246	...	3119
	Regulate flow	8	2274	246	...	3162
	Swing	7	2274	1241	...	3014
	Photograph	5	2274	4992	...	2496
	Clean	5	2274	4992	...	2497
	Provide acceleration	7	2274	22	...	2600
	Minimize noise	7	2274	16	...	3145

**Figure 12.** Shortest path retrieval results for the “drive-open shell.”

Then, the sum of the out-degree and in-degree of each node is calculated, respectively, which are labeled as follows: “adapt to diameter (5), drainage pipeline robot (16), connect (239), auxiliary shaft (4), bite (4), pliers (4), crocodile (3), mechanical crocodile (2), swim (6), shark (2), mechanical shark (4),

submarine (5), Archimedes principle (1).” Based on Eqs. (5 and 6) in step C4, the absolute domain distance of “adapt to diameter” is calculated as $D_{\text{adapt to diameter}}^{\text{adapt to diameter}} = 12/13 \times (1/5 + 1/16 + 1/239 + \dots + 1/5 + 1) = 3.6616$. Then, the domain distance $D_{\text{adapt to diameter}}^{\text{adapt to diameter}}$ is calculated as $D_{\text{adapt to diameter}}^{\text{adapt to diameter}} = 3.6616/D_{\text{max}}$

Table 11. The domain distance calculation results for the F_2 's function analogical sources

Search term	Function analogical sources	Hop count		Degree sum ($\sigma_k^t + \bar{r}_k^t$)			Domain distance
Drive	Open shell	12	54	28	...	1	1.0000
	Worm	10	54	28	...	1	0.8633
	Slide	11	54	16	...	1	0.8243
	Swim	11	54	16	...	1	0.7612
	Control buoyancy	11	54	16	...	2	0.6562
	Rolling support parts	8	54	34	...	1	0.6120
	Fly	7	54	17	...	1	0.6007
	Transfer motion	10	54	34	...	1	0.5738
	Directional transfer	8	54	2	...	1	0.5716
	Cut vibration	6	54	2	...	1	0.5123

Table 12. AV calculation results for the F_2 's function analogical sources

Search term	Function analogical sources	Domain distance	$S_{(F, F')}$	AV
Drive	Open shell	1	0.5039	0.7099
	Control buoyancy	0.6562	0.7174	0.6861
	Slide	0.8243	0.5286	0.6601
	Rolling support parts	0.6120	0.6886	0.6492
	Worm	0.8633	0.4800	0.6437
	Swim	0.7612	0.5124	0.6245
	Directional transfer	0.5716	0.6094	0.5902
	Regulate flow and speed	0.4455	0.7515	0.5786
	Cut vibration	0.5123	0.6489	0.5766
	Connect	0.4461	0.6479	0.5376

= 1. Similarly, the domain distances of the other effect analogical sources are calculated. Due to the limitation of the article length, only the top 10 effects' analogical source domain distances are shown in Table 14.

According to the effect analogical sources retrieved from Table 14, use the MATCH function (e.g., MATCH (n1: effect {name: "Archimedes principle"})-[: Achieved_by*1..5]-(p1: function) RETURN p1.) to retrieve the relevant functions of each effect node. The retrieval results are shown in Figure 15.

According to steps D1–D3, the semantic similarity $S_{(F, E)}$ between effect words and search terms in Table 14 is calculated using Eq. (7). For example, the semantic similarity between "adapt to diameter" and "Archimedes principle" is calculated as $S_{(F, E)}^{\text{name}} = \text{sim}(\mathbf{u}_F, \mathbf{u}_E) = 0.5945$. Then, the function nodes with the "Achieved_by" relationship for the "Archimedes principle" effect are searched for "control buoyancy," and the semantic similarity between the "adaptive diameter" and "control buoyancy" functions are calculated as $S_{(F, F')}^{\text{function}} = \text{sim}(\mathbf{u}_F, \mathbf{u}_{F'}) = 0.6064$. Similarly, the similarity between the search term and other effect analogical sources is calculated. Considering the article length limitation, only the top 10 effect analogical sources with similarity are shown in Table 15.

For example, AV calculations for "ADAPT to diameter-Archimedes principle." Eq. (9) is used to calculate the AV between the search term and the effect analogical source, $AV(\text{Archimedes principle, adapt to diameter}) = \sqrt{0.6005 \times 1} = 0.7749$. The calculation results of the top 10 effect analogical sources with AV are shown in Table 16.

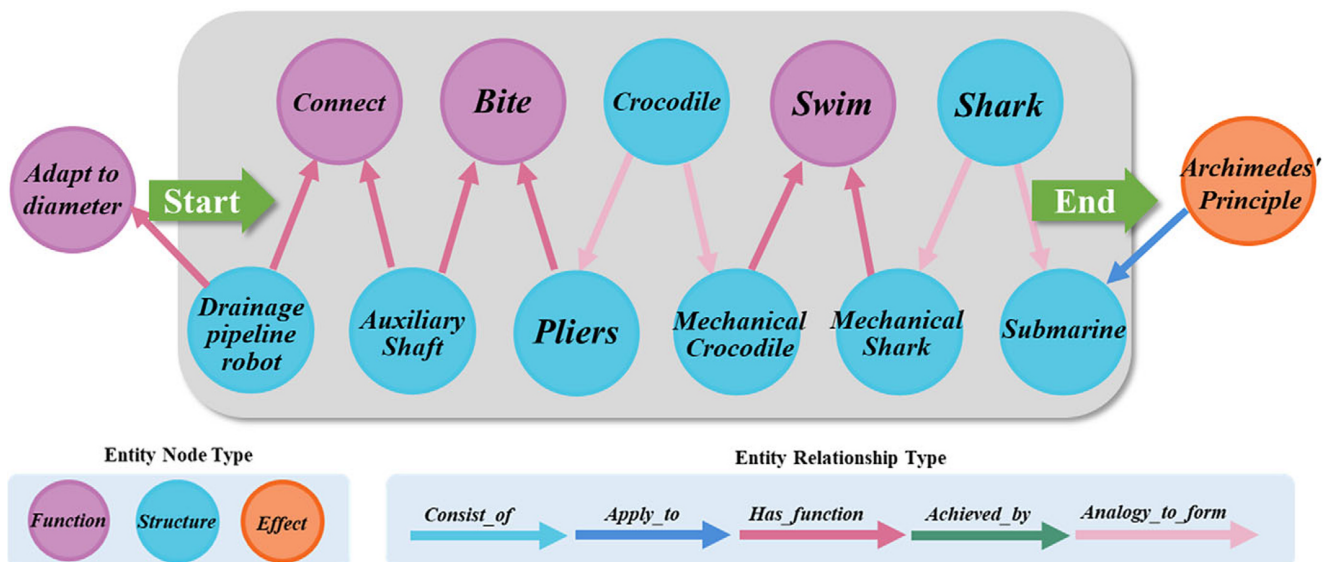
From Table 16, the top five effect analogical sources (marked in bold in Table 16) are selected for the solution of the search term "adapt to diameter," including Archimedes principle (0.7749), Snake's winding effect (0.6905), vibration effect (0.5488), friction effect (0.4905), and pressure sensing principle (0.4582). The properties of five function nodes are displayed, as shown in Figure 16. For example, the ID of the effect "Archimedes principle" node is 3493, the scientific laws is "Hydrostatic Equilibrium," the phenomenon description is "An object immersed in a fluid will experience upward buoyancy," the effect type is physical effects, the input flow is "volume," and the output flow is "power."

Node attribute	Node attribute	Node attribute	Node attribute	Node attribute
Function	Function	Function	Function	Function
<id> 3203	<id> 3361	<id> 3411	<id> 3156	<id> 3336
Name control buoyancy	Name slide	Name worm	Name rolling support parts	Name swim
Fv control	Fv slide	Fv worm	Fv rolling support	Fv swim
Fn buoyancy	Fn N/A	Fn N/A	Fn parts	Fn N/A

Figure 13. Top five function node attributes ranked by AV (translated from Chinese into English).

Table 13. Search results for the shortest path nodes of F_6 in DbAKG

Search term	Effect analogical sources	Node count	Node ID of the shortest path			
			Node ₁	Node ₂	...	Node _n
Adapt to diameter	Archimedes principle	13	2328	80	...	3493
	Snakes winding	7	2328	16	...	4861
	Vibration effect	7	2328	4252	...	3463
	Friction effect	7	2328	16	...	3503
	Principle of pressure sensing	7	2328	4252	...	3473
	Dielectric polarization	6	2328	4252	...	3515
	Thermal expansion	6	2328	16	...	3464
	Thermionic emission	7	2328	16	...	3513
	Bimetallic strip effect	7	2328	16	...	3434
	Thermal radiation	7	2328	16	...	3448
	Faraday's law of induction	6	2328	16	...	3529
	Second-order phase transition	5	2328	16	...	3484
	Vacuum effect	6	2328	16	...	3599
	Joule–Thomson effect	7	2328	4252	...	3474
	Gas compression	6	2328	16	...	3554
	Photovoltaic effect	5	2328	16	...	3487
	Energy dissipation effect	6	2328	16	...	3537
	Magnetic field lines	5	2328	22	...	3520
	Magnetic field strength	5	2328	22	...	3618
	Inductance	6	2328	16	...	3512

**Figure 14.** Shortest path retrieval results for the “adapt to diameter-Archimedes principle.”

Innovative design of PIR based on analogical transfer

The example of F-F analogical transfer: a worm PIR

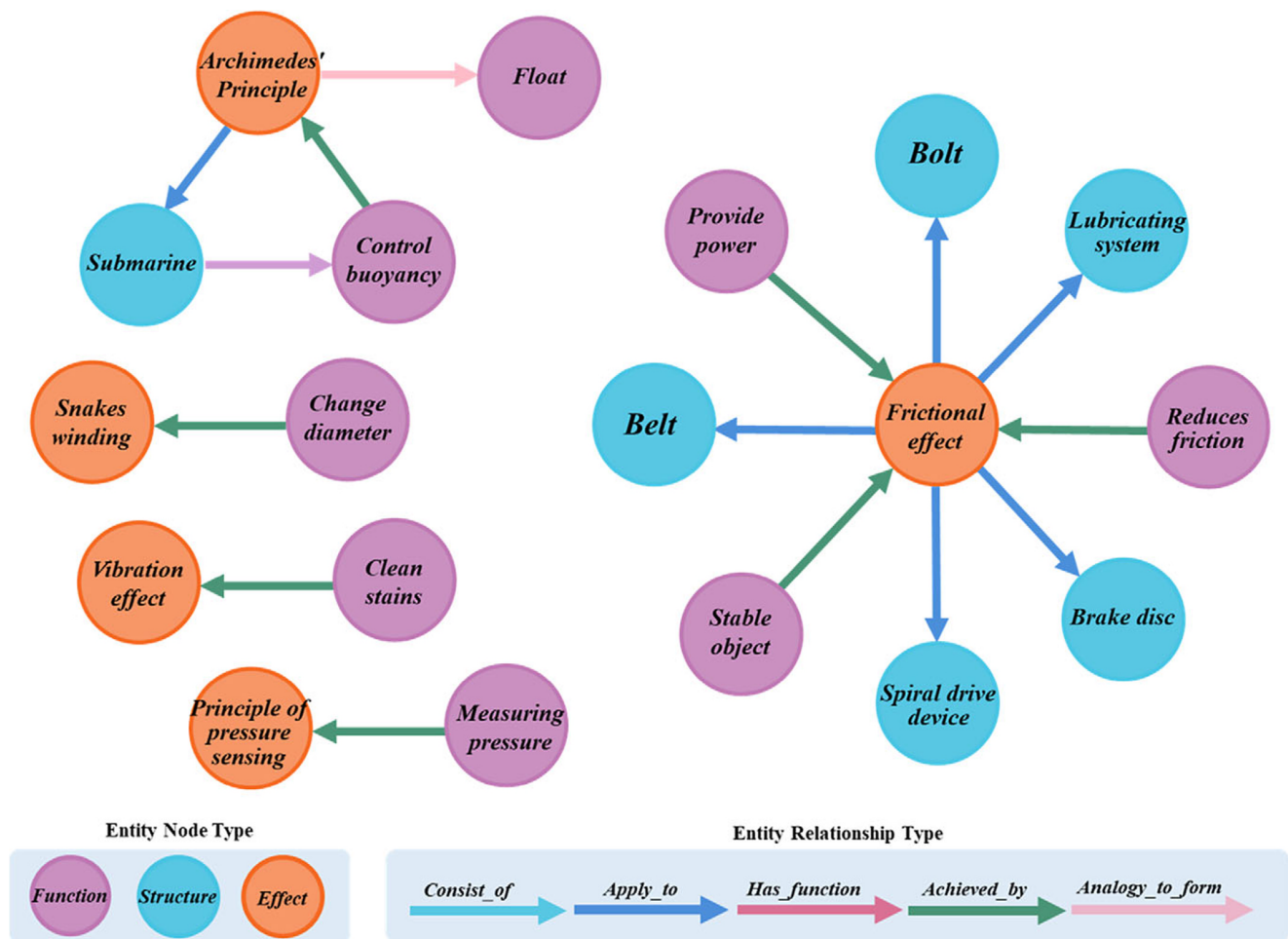
Select the search term “drive” and choose the results with high AV analogical sources, namely “control buoyancy,” “slide,” “worm,”

“rolling support parts,” and “swim,” which are used to support the function analogical transfer.

Based on step E1, combined with the commonly used PIR driven walking method, use the search results of the analogical source to transfer Fv. Derive “roll” from “rolling and supporting parts,” and

Table 14. The domain distance calculation results for the F_g 's effect analogical sources

Search term	Effect analogical sources	Hop count	Degree sum ($\sigma_k^t + f_k^t$)				Domain distance
Adapt to diameter	Archimedes principle	12	5	16	...	1	1.0000
	Snakes winding	6	5	133	...	1	0.6461
	Vibration effect	6	5	21	...	1	0.4620
	Friction effect	6	5	133	...	8	0.3917
	Principle of pressure sensing	6	5	21	...	2	0.3213
	Dielectric polarization	5	5	21	...	2	0.3156
	Thermal expansion	5	5	133	...	3	0.3005
	Thermionic emission	6	5	133	...	2	0.2958
	Bimetallic strip effect	6	5	133	...	3	0.2794
	Thermal radiation	6	5	133	...	3	0.2794

**Figure 15.** Search results for functions related to effects.

obtain the transfer words “walk” from “control buoyancy” and “swim,” leading to “crawl.”

Based on step E2, the scene is transferred, and “roll” can be associated with the “driving wheel” movement in the design of PIR. The “crawl” transfers to the working scene of the robot inside the pipeline, which requires the robot to have a high degree of flexibility

and adaptability, considering the closed and irregular nature of the pipeline.

Based on step E3, combining the function “drive” and “crawl,” a novel worm-PIR has been designed, capable of propelling itself within pipelines through body contractions and extensions. The peristaltic motion of the robot not only provides the necessary

Table 15. Similarity calculation results for the F_6 's effect analogical sources

Search term	Effect analogical sources	$S_{(F,E)}^{name}$	Related functions	$S_{(F,F')}^{function}$	$S_{(F,E)}$
Adapt to diameter	Archimedes principle	0.5945	Control buoyancy	0.6064	0.6005
	Snake's winding	0.5553	Change diameter	0.9204	0.7379
	Friction effect	0.6450	Transmit power, reduce friction, stabilize object	0.6590	0.6520
	Vibration effect	0.5355	Clean stains	0.6930	0.6143
	Principle of pressure sensing	0.6478	Measure pressure	0.6590	0.6534
	Dielectric polarization	0.5759	Store charge	0.5892	0.5826
	Thermal expansion	0.6197	Generate work	0.4987	0.5592
	Thermionic emission	0.6544	Lower temperature	0.6997	0.6771
	Bimetallic strip effect	0.6257	Measure temperature	0.6648	0.6453
	Thermal radiation	0.5694	Measure temperature	0.6648	0.6171

Table 16. AV calculation results for the F_6 's effect analogical sources

Search term	Effect analogical sources	$S_{(F,E)}$	Domain distance	AV
Adapt to diameter	Archimedes principle	0.6005	1	0.7749
	Snake's winding	0.7379	0.6461	0.6905
	Vibration effect	0.652	0.462	0.5488
	Friction effect	0.6143	0.3917	0.4905
	Pressure sensing principle	0.6534	0.3213	0.4582
	Thermionic emission	0.6771	0.2958	0.4475
	Second-order phase transition	0.654	0.2764	0.4252
	Bimetallic strip effect	0.6453	0.2794	0.4246
	Joule–Thomson effect	0.6347	0.2744	0.4173
	Faraday's law of induction	0.6269	0.2777	0.4172

propulsion but also allows for flexible steering within the pipeline, adapting to various pipeline environments.

After that, a worm PIR CS is constructed by flexible rolling structure, axial expansion structure, cardan joint, and radial extension worm structure, as shown in Figure 17. However, there is a physical contradiction in the scheme, that is, the direction of the ordinary driving wheel can only be simply forward and backward, and the attitude cannot be adjusted in time. To solve this problem, the force is decomposed into orthogonal radial force and tangential force, and the driving force is divided into orthogonal radial force and tangential force. Subsequently, a mecamum wheel is used to output the final concept, and the wheel drive module and creep mechanism work together to achieve both rolling and creep motion, which can be detected inside the pipe of different diameter sizes. At the same time, the scheme reduces mechanical wear and prolongs the service life of the robot by imitating the movement pattern of snakes in nature. The example declared a China Patent (CN117489912A) to support the validation of the feasibility of F-F analogical transfer in stimulating design innovation.

The example of F-E analogical transfer: A snake PIR

Select “adaptive diameter” as the search term, and select the analogical sources with high AV, namely “snake's winding,” “Archimedes principle,” “friction effect,” “vibration effect,” and “pressure sensing principle,” to effectively support the transfer.

Based on step F1, characteristic extraction is carried out for the selected effects. Taking the effect “snake's winding” as an example, the phenomenon description is to adapt the diameter of the cylinder. The difference is that in the snakes climbing trees, the snakes adapt to the outside diameter of the trunk, while the PIR needs to adapt to the inside diameter of the pipeline. The input and output flows of “snake's winding” are “biological energy” and “kinetic energy,” respectively, mapped to “electrical energy” and “kinetic energy” in PIR design. This process inspires the possibility of developing PIRs with a serpentine structure.

Based on step F2, the physical parameters of natural phenomena are abstracted into upper-level words and mapped into function components in robot design. For example, (a) the flexibility of the snake body is abstracted as a “multidegree of freedom and scalable structure” and transferred to the spiral drive and axial expansion structure of the robot. (b) The buoyancy of the object is abstracted as “underwater buoyancy control” and mapped to the buoyancy propulsion function. (c) The friction coefficient of the snake scale is abstracted as a “friction driving unit” and mapped to the rubber friction surface. (d) The vibration frequency of the snake's body is abstracted as a “vibration cleaning system” and mapped to vibration frequency regulation.

Based on step F3, the application examples of these effects are decomposed. The snake body is transformed into a spiral driving unit to simulate its spiral advance and pipeline navigation. The object's buoyancy control is transformed into a buoyancy propulsion unit to simulate underwater dynamics and stability. The contact part of the snake scale is transformed into a friction driving unit to simulate the adhesion and propulsion of the pipe wall. The vibration of the snake body is converted into a vibration cleaning unit that simulates the cleaning mechanism of the inner wall of the pipe.

Based on the aforementioned work, a new CS of snake PIR is described in Figure 18. The initial proposal faced two major contradictions: (1) a single movement mode could not adapt to the complex and changing environmental requirements and (2) a simple support structure could not cope with the challenges posed by the changing diameter of the pipeline. Regarding contradiction 1, the time separation strategy is adopted to execute two different

Node attribute	Node attribute	Node attribute
Effect	Effect	Effect
<id> 3493	<id> 4861	<id> 3463
Name Archimedes' Principle	Name Snake's winding	Name Friction effect
Types Physical effects	Types Biological effects	Types Physical effects
Scientific laws Hydrostatic Equilibrium	Scientific laws Coulomb's law of friction	Scientific laws Coulomb's law of friction
Phenomenon description An object immersed in a fluid will experience ...	Phenomenon description Snakes adjust the contact area between their body ...	Phenomenon description The frictional effect describes the resistance of two contact...
Input flow Volume	Input flow Bioenergy	Input flow Contact pressure
Output flow Buoyancy	Output flow Kinetic energy and potential energy	Output flow Resistance force

Node attribute	Node attribute
Effect	Effect
<id> 3503	<id> 3473
Name Vibration Effect	Name Pressure Sensing Principle
Types Physical effects	Types Physical effects
Scientific laws Law of harmonic motion	Scientific laws Piezoelectric law
Phenomenon description Vibration Effect involves the oscillatory motion of...	Phenomenon description The Pressure Sensing Principle describes how pressure sensors...
Input flow Excitation Force	Input flow Mechanical Pressure
Output flow Vibration Response	Output flow Electrical Signal

Figure 16. Top five effect node attributes ranked by AV (translated from Chinese into English).

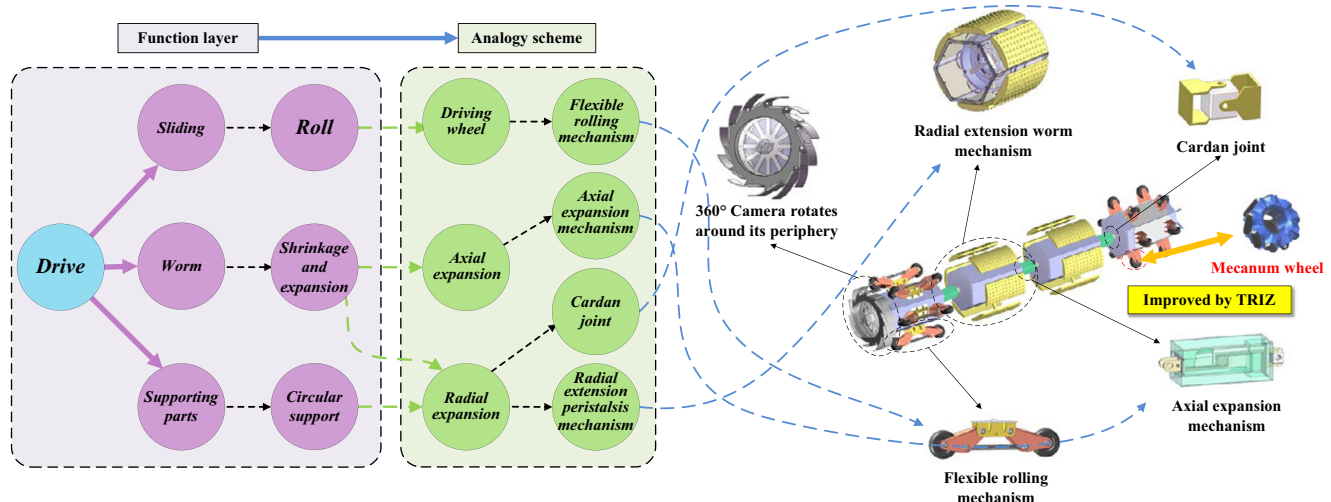


Figure 17. A worm PIR-based analogical design scheme.

movement modes simultaneously. For terrain with silt, a spiral movement is adopted, while for others, a wheeled movement is used. Regarding contradiction 2, a modular design is adopted to accommodate pipelines of different diameters by assembling different modules. This serpentine PIR design combines two movement modes to adapt to diverse environments and enables adaptive support for different pipeline diameters through a split design. In addition, the vertical cross-connect technology of the rack and pinion enables efficient adaptive turning. The design has been

applied for China Patent (CN116105009A) to verify the effectiveness of F-E analogical transfer in stimulating design innovation.

To improve the efficiency of analogical design and manage design knowledge effectively, KG-AAD prototype system is developed based on Neo4j graph database and MySQL database. The system consists of three core function modules: (1) knowledge entity management, (2) analogical source retrieval, and (3) analogical transfer, as shown in Figure 19. These modules help designers obtain suitable analogical sources, simplify the characteristic

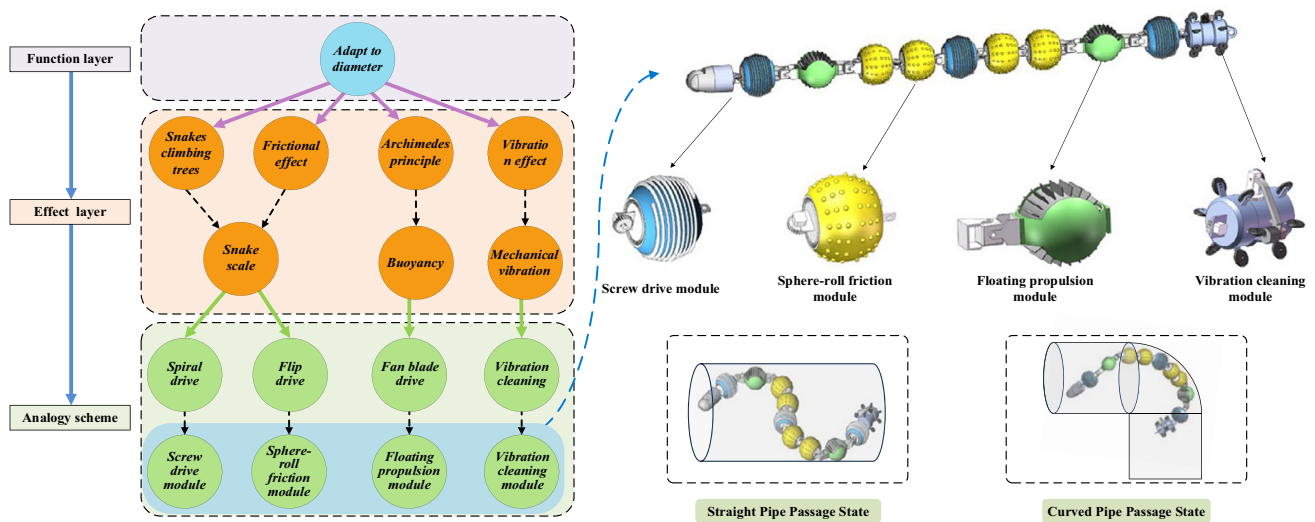


Figure 18. A snake PIR-based analogical design scheme.

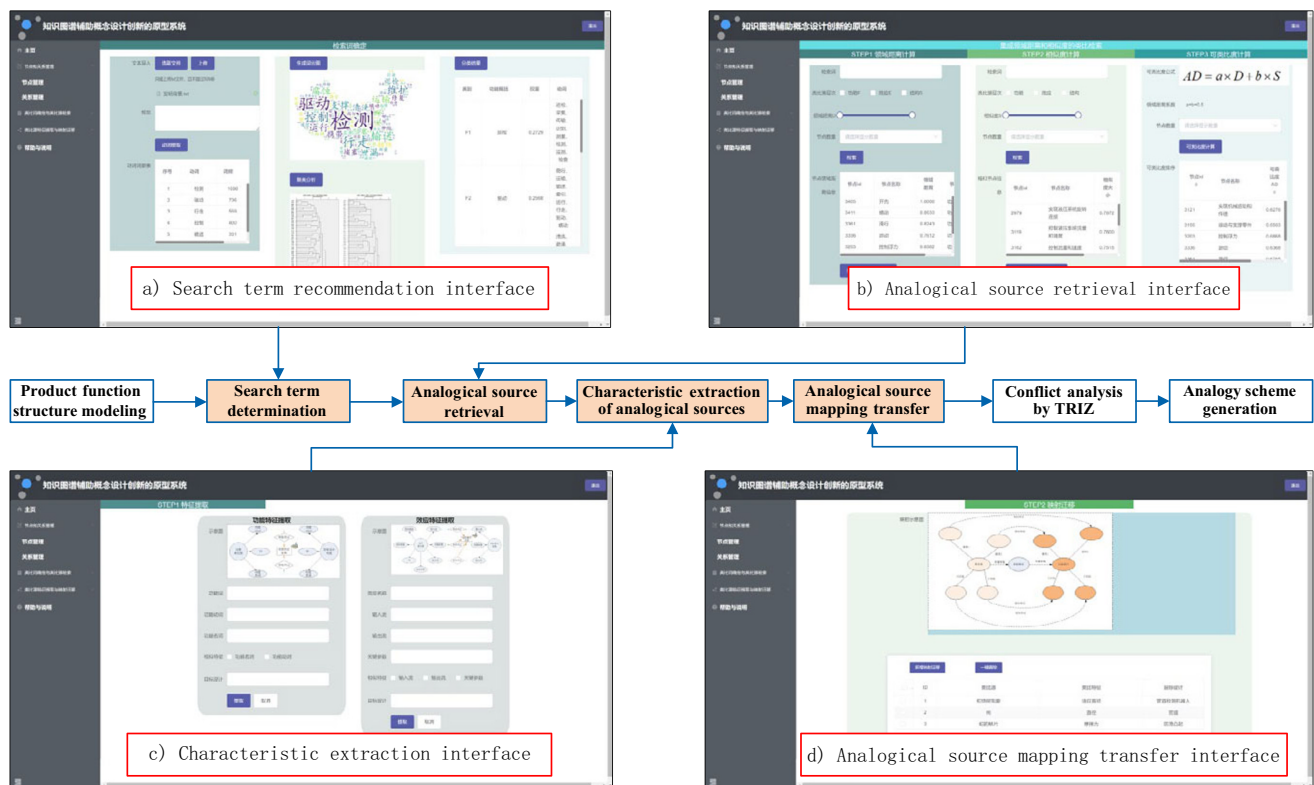


Figure 19. KG-AAD prototype system.

extraction and rapid transfer process, ensure that the selected analogical source is highly relevant to the design target, and tap its potential innovation value, thereby broadening the design vision and enhancing the innovation effect.

Comparison of DbA design schemes and discussion of results

To verify the advantages of the analogy method and the generated CS in this study, this section will compare and analyze the two new

CSs obtained by the analogical design with the existing invention patents of PIR. Refer to the evaluation model proposed by Hao et al. (2017), a cosine distance model based on the weighted average word vector of word2vec is constructed to calculate the differences among different schemes, as shown in Eqs. (10–12).

$$Novelty(pc) = \frac{1}{p} \sum_{s=1}^p dis(pc, ec_s) \quad (10)$$

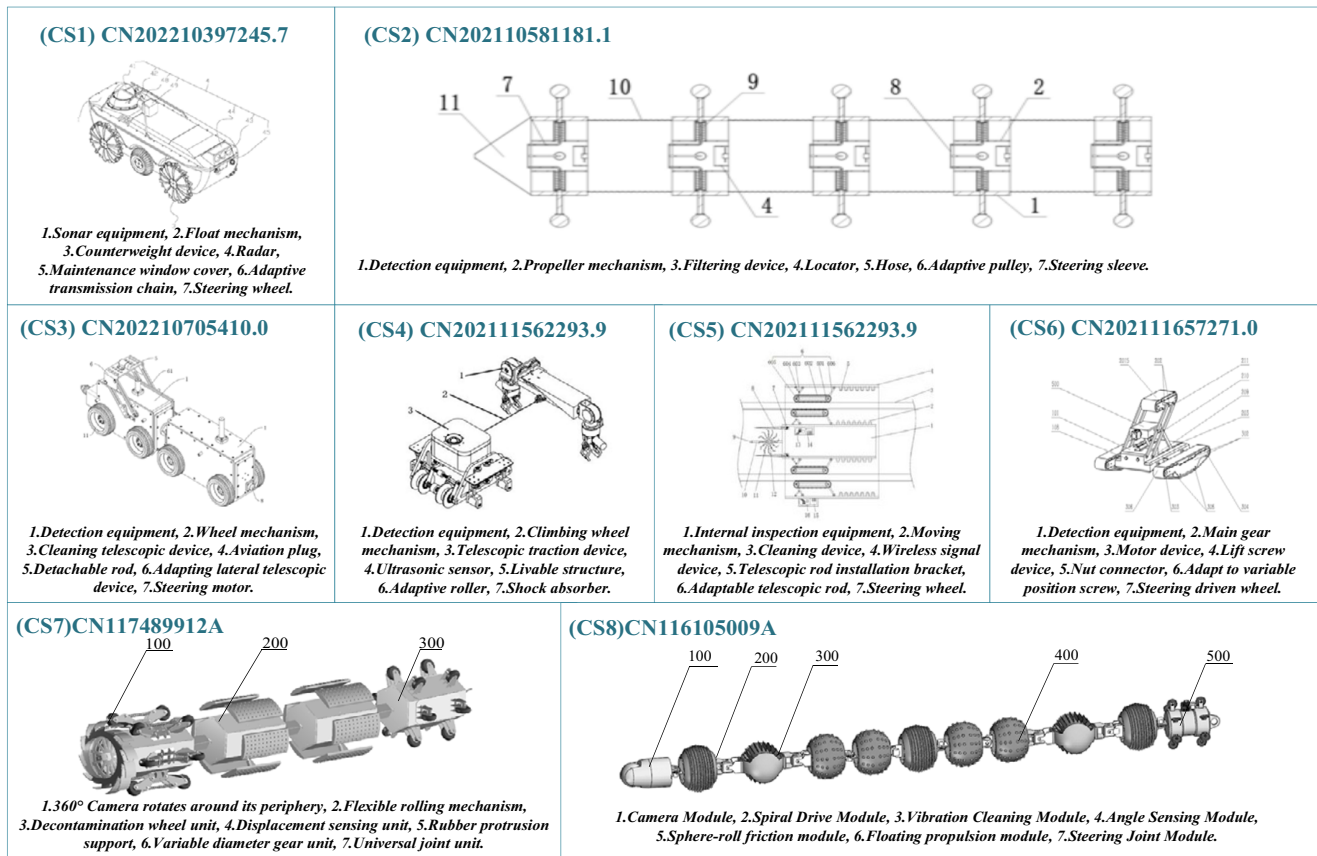


Figure 20. Abstract figures of eight patents: (CS1) PIR, (CS2) a serpentine PIR, (CS3) dual-drive PIR, (CS4) a separable PIR, (CS5) a cleaning PIR, (CS6) A tracked PIR, (CS7) a worm PIR, and (CS8) a snake PIR.

$$dis(pc, ec) = \frac{\mathbf{u}_{pc} \cdot \mathbf{u}_{ec}}{|\mathbf{u}_{pc}| \times |\mathbf{u}_{ec}|} \quad (11)$$

$$\mathbf{u}_{pc} = \sum_{i=1}^q \alpha_i \mathbf{u}_i \quad (12)$$

where $Novelty(pc)$ represents the average distance between the current evaluated scheme pc and ec_s , that is, the novelty of the CS, and $1 \leq s \leq p$; $dis()$ is the distance between the weighted average word vectors \mathbf{u}_{pc} and \mathbf{u}_{ec} of the two CSs; α_i represents the weight of the function of the principle solution (PS) in Section “Retrieval of analogical sources for PIR innovation design,” and \mathbf{u}_i is the i th PS’s word vector in the CS pc , and $1 \leq i \leq q$.

Then, to evaluate the applicability of the CS or PS, Eq. (11) is used to calculate the distance between PSs under the same CS, as shown in Eq. (13).

$$Feasibility(pc) = 1 - \frac{2}{F(F-1)} \sum_{i=1}^F \sum_{j=i+1}^F dis(f_i, f_j) \quad (13)$$

where $Feasibility(pc)$ indicates the feasibility of the design project pc , f_i and f_j are the i th and j th PSs of pc , respectively, and F is the total number of PSs of pc .

Similarly, using F_2 (drive) and F_6 (adapt to diameter) as search terms, set search rules in [innojoy.com](https://www.innojoy.com) and output the top 6 relevant authorized patents, and compare the six schemes (named CS1–CS6) with the tow CSs (named CS7 and CS8), as shown in Figure 20.

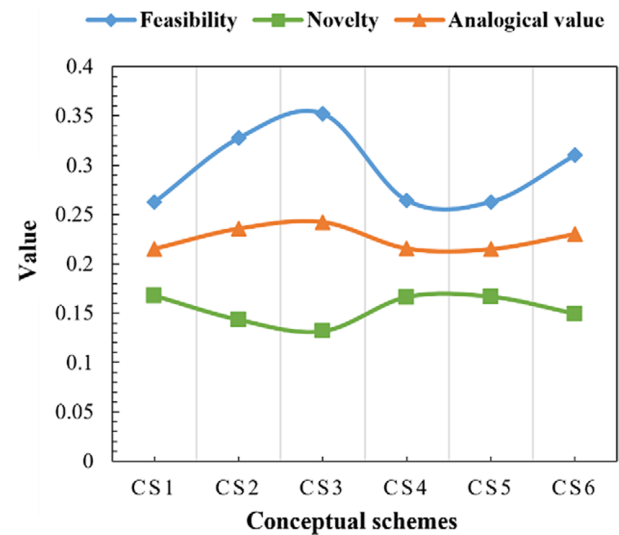


Figure 21. Trend analysis of AV, novelty, and feasibility scores for six CSs.

In addition, seven key functions of each scheme (derived from the function clustering results in Section “Retrieval of analogical sources for PIR innovation design”) and their PSs are selected to facilitate the comparison of novelty and feasibility among the eight schemes.

After that, the settlement results of novelty, feasibility, and AV of CS1–CS6 are described in Figure 21. It can be seen that CS4 and CS5

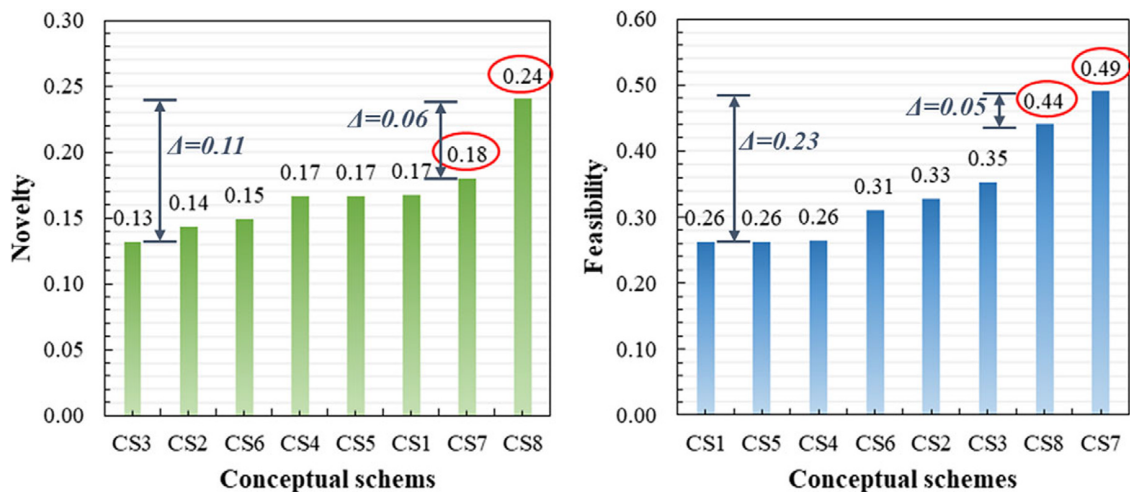


Figure 22. Novelty and feasibility evaluation scores of eight CSs.

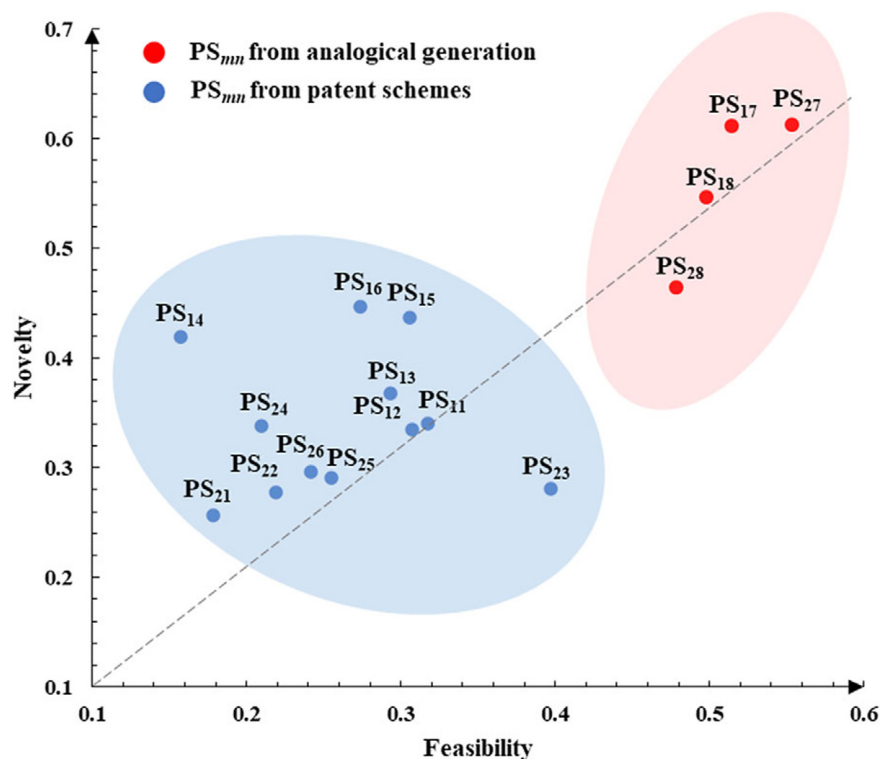


Figure 23. Distribution of novelty and feasibility scores in “drive” and “adapt to diameter” function PSs.

score high in novelty but have unsatisfactory performance in adaptability. For example, CS4 had a novelty score of 0.17, ranking first among the 6 CSs, but its feasibility score is only 0.26, ranking sixth among the 6 CSs. Compared with CS4, CS3 ranked first in the feasibility score of 0.35, but its novelty score is only 0.13 and ranked sixth. It is evident that novelty and feasibility are often used to evaluate the single performance of a CS, but it is difficult to provide a comprehensive perspective to evaluate the CS. Compared with $Novelty(pc)$ and $Feasibility(pc)$, it can be intuitively seen that a CS with high AV takes both novelty and feasibility into consideration, so that its product can obtain higher user satisfaction after being put

into the market. In addition, AV can help identify those CSs that do not perform well on a single indicator, but perform well in the comprehensive evaluation, and describe the design potential and practical value of the CS from the perspective of the overall design value.

Then, the novelty and feasibility of eight CSs are calculated by Eqs. (10–13), and the novelty scores and feasibility scores of each CS are ranked from low to high, respectively, as shown in Figure 22. The novelty score ranking of the CS is $CS8 > CS7 > CS1 = CS5 = CS4 > CS6 > CS2 > CS3$, and the feasibility score ranking of the CS is $CS7 > CS8 > CS3 > CS2 > CS6 > CS4 = CS5 = CS1$.

Table 17. Comparison with other relevant literature on analogical design methods

Author (year)	Topics				
	Design knowledge representation	Data sources	Analogy retrieval calculation model	Analogical transfer strategy	Application tool
Fu et al. (2015)	Function	Patents	Function similarity, function domain distance	N/A	N/A
Srinivasan et al. (2018)	Conceptual principle scheme	Design project data	Project domain distance technology	N/A	N/A
Jia et al. (2019)	Refined SBF	Case base	Function similarity	N/A	N/A
Song and Fu (2019)	Component behavior material	Patents	Similarity of components, behaviors, and materials	N/A	N/A
Chen et al. (Chan et al., 2021)	Function	Patents	Function similarity	N/A	Function converter
Cao et al. (2021)	FBS	Patents and websites	Fuzzy triangular numbers-based	Functional innovation strategies in biological analogies	N/A
Song and Fu (2022)	Function-behavior-principle-component	Patents	Function similarity	N/A	An analogical tool (VISION)
Nie et al. (2022)	Function-structure effect-parameters	Patents, case base	Decomposed function similarity	TRIZ	Design process model based on digital twins
Liu et al. (2023a)	SMFM	Patents	Function similarity, meta-function similarity	A second iterative analogical strategy from function to meta-function	Analogical reasoning-based computer aided conceptual design tool
Chen et al. (2024)	Source-function-application	Websites	Problem-driven similarity solution-driven similarity	Application-function-source link, source-function-application link	AksNatureNet
Proposed approach	I-FES	Patents, websites, case base	F-F and F-E similarity, F-F and F-E domain distance	Function-driven attribute mapping analogical transfer strategy, TRIZ	KG-AAD system

It can be seen that CS8 and CS7 both have higher scores in terms of novelty and feasibility. The optimal CS8 has greater advantages in feasibility and novelty compared with the worst scheme, such as $\Delta_{feasibility}(CS8-CS4) = 0.23$, and $\Delta_{novelty}(CS8-CS3) = 0.11$. Meanwhile, CS8 has a certain advantage in novelty compared with CS7, and the dominant degree of novelty and adaptability $\Delta_{novelty} = 0.06 > \Delta_{feasibility} = 0.05$.

In DbA, considering the effects of snake crawling and its application in the external biological field infuses innovative ideas into design thinking. This design approach is helpful to break the traditional mindset of relying solely on the mechanical or control domain, driving a whole new way of thinking about the design of PIR. The resulting serpentine PIR design can be adapted to a diverse pipeline environment, showing a broader market potential.

To further verify the advantages of the proposed approach, the PSs of F_2 and F_6 of the eight CSs are taken as examples and named as PS_{mn} in order (from Figure 20), where $m \in \{1, 2\}$ correspond to F_2 and F_6 , respectively, n represents the eight schemes. According to Eqs. (10–13), the novelty and feasibility of the PS of the function “drive” and function “adapt to diameter” in each CS are calculated, as shown in Figure 23. By comparing the scores of novelty and feasibility, it is found that the closer the PS is to the upper right corner, the higher the score of the two indexes, which emphasizes the superior comprehensive design value of the PSs. In addition, it can be seen from Figure 23 that the four PSs obtained by the

analogical design (red area in the figure), namely PS_{17} , PS_{18} , PS_{27} , and PS_{28} . The novelty and feasibility scores of PS_{17} (0.51, 0.61), PS_{18} (0.50, 0.55), PS_{27} (0.55, 0.61), and PS_{28} (0.48, 0.46) are concentrated near the diagonal, respectively. It shows that the analogy method has significant innovation advantages and potential technical value in driving the conceptual design of product innovation.

Through comparison with other relevant literature related to DbA, the design knowledge representation, data source, analogical retrieval calculation model, and analogical transfer strategy are compared and discussed, indicating that KG provides a new search approach and knowledge transfer strategy method for analogical design, as shown in Table 17.

- First, the proposed I-FES model takes into account the effects of knowledge to inspire design innovation and enrich the entity attributes, which can capture the design principle and increase the innovation possibility of analogical sources.
- Second, this article uses multisource design information, such as patent text, TRIZ effect webpage, and historical case base to ensure the richness and reliability of analogical knowledge sources.
- Third, this study considers the domain distance and semantic similarity of F-F and F-E and considers the practicality of analogical source while focusing on novelty, which enhances the accuracy of analogical retrieval.
- Fourth, most of the work neglects the formulation of analogical transfer strategy and the effect of knowledge transfer is subjective.

This article proposes that the attribute matching rule of function provides an objective solution strategy for analogical transfer.

Conclusion and future work

To address the challenges designers face in uniformly representing and utilizing analogical knowledge in the existing DbA process, as well as the difficulty in obtaining novel and applicable analogical sources during analogical source retrieval, a DbAKG-driven product innovation design approach is proposed. First, the I-FES analogical knowledge ontology is proposed to construct the DbAKG. Second, the AV model is constructed to quantify the analogical source of retrieval by integrating knowledge domain distance and semantic similarity. Furthermore, F-F and F-E are constructed to deduce the new schemes by analogy, and TRIZ conflict resolution strategies are used to improve the technical conflicts of the scheme. Finally, taking the “drive” and “adapt to diameter” functions of PIR as an example, the worm PIR and snake PIR designs are generated by using F-F and F-E analogical transfer strategies, effectively verifying the feasibility of the proposed approach. In addition, the KG-AAD prototype system is developed to provide a computer-aided design tool for product innovative design.

Compared with traditional analogical design, which mainly focuses on structure analogy process, this study explores and utilizes the semantic relations and matching rules of different design knowledge from the perspective of horizontal and vertical transfer of knowledge levels, and provides a new solution framework for analogical design by using KG, and promote the cross-domain inspiration of analogical design thinking. Based on the developed DbAKG, the AV is constructed to retrieve the multidomain analogical sources, which has three research priorities:

- The constructed DbAKG provides a unified representation of function, effect, structure, and their analogical relationships, stores 5579 knowledge nodes and 7335 relationships, and expands the breadth and depth of analogical sources.
- The AV model of integration with domain distance and semantic similarity is constructed, which provides an index for the correlative path retrieval of analogical sources in DbAKG, and the KG-AAD system is developed to improve the efficiency of DbA.
- Based on DbAKG, F-F and F-E analogical transfer strategies are constructed. Taking PIR design as an example, an analogical CS incorporating cross-domain knowledge is generated, which solves the problem of traditional analogical transfer relying on design experience and subjective preference.

All scientific studies have limitations, and this work is not an exception. The limitations are summarized as follows. (1) The proposed analogical design focuses more on dealing with design knowledge with direct or indirect semantic relations to provide new design ideas, ignoring the design potential contained in the implicit relations between knowledge. (2) The initial design scheme driven by DbAKG has the characteristics of diversity and multidomain, and the previous multicriteria evaluation model cannot truly capture the potential value of cross-domain schemes. (3) The proposed entity extraction rule is a fixed extraction based on semantic dependency and does not involve dependency syntax analysis of specific sentences, which reduces the possibility of new knowledge entities being mined.

Future research can focus on the following directions:

- Semantic similarity is used to explore the implicit relationship of design knowledge, and physiological signals such as EEG and eye movement are used to capture the inspiring effect of potential analogical knowledge on design thinking, and further enrich the DbAKG.
- A scheme similarity evaluation method based on graph neural network model is established to explore the internal correlation of different schemes from the semantic perspective of graph embedding and to achieve rapid clustering and value screening of initial schemes.
- Generative language models are used to explore new entity extraction rules, while physiological signals (such as EEG and eye tracking) provide real-time feedback on the design thinking process, capturing valuable insights to inspire analogical design.

Abbreviations

AV	analogical value
C-K	concept–knowledge
CS	conceptual scheme
DbA	design-by-analogy
DbAKG	design-by-analogy knowledge graph
DR	dependency relation
FBS	function-behavior-structure
Fn	function noun
Fv	function verb
I-FES	improved function-effect-structure
KG-AAD	knowledge graph-assisted analogical design
LTP	language technology platform
NLP	natural language processing
PIR	pipeline inspection robot
PS	principle solution
SMFM	structure-mapping function model

Data availability statement. No data were used for the research described in the article.

Author contribution. **Liting Jing:** Conceptualization and methodology. **Yubo Dou:** Formal analysis. **Qizhi Li:** Writing – original draft. **Di Feng:** Investigation. **Mingyang Huang** and **Di Feng:** Writing – review and editing. **Liting Jing** and **Shaofei Jiang:** Funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding statement. This work was supported by the National Natural Science Foundation of China (grant number 52105282), the Zhejiang Provincial Natural Science Foundation of China (grant number LQ22E050011), the China Postdoctoral Science Foundation (under grant number 2024M752862), the Key Research and Development Program of Zhejiang Province (grant number 2024C01236), and the Jinyun Innovation Design Institute Project of Zhejiang University of Technology (grant number SKY-HX-20220296). These sources of support are gratefully acknowledged.

Competing interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

References

- Avdeenko TV and Makarova ES (2018) Knowledge representation model based on case-based reasoning and the domain ontology: Application to the IT consultation. *IFAC PAPERSONLINE* 51, 1218–1223.
- Beitz W, Pahl G and Grote K (1996) Engineering design: A systematic approach. *MRS Bulletin* 71, 30.

- Bhattacharya K, Majumder A, Bhatt AN, Keshwani S, Venkataraman S and Chakrabarti A (2024) Developing a method for creating structured representations of working of systems from natural language descriptions using the SAPPhIRE model of causality. *AI EDAM* 38, e24.
- Cao G, Sun Y, Tan R, Zhang J and Liu W (2021) A function-oriented biologically analogical approach for constructing the design concept of smart product in industry 4.0. *Advanced Engineering Informatics* 49, 101352.
- Casakin H, Ball LJ, Christensen BT and Badke-Schaub P (2015) How do analogizing and mental simulation influence team dynamics in innovative product design? *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 29, 173–183.
- Chan J, Fu K, Schunn C, Cagan J, Wood K and Kotovsky K (2011) On the benefits and pitfalls of analogies for innovative design: Ideation performance based on analogical distance, commonness, and modality of examples. *Journal of Mechanical Design* 133(8), 081004.
- Chan EM, Kor AL, Ng KW, Ang MC and Wahab ANA (2021) A conceptual design framework based on TRIZ scientific effects and patent mining. *International Journal of Advanced Computer Science and Applications* 12(12), 43–50.
- Chen W, Li Z and Liu T (2010) LTP: A chinese language technology platform. *Coling 2010: Demonstrations* 2010(08), 13–16.
- Chen L, Cai Z, Jiang Z, Luo J, Sun L, Childs P and Zuo H (2024) AskNature-Net: A divergent thinking tool based on bio-inspired design knowledge. *Advanced Engineering Informatics* 62, 102593.
- Chen C, Li Y, Tao Y, Chen J, Liu Q and Li S (2020) A method to automatically push keywords for biological information searching in bio-inspired design. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 235, 30–47.
- Chou J-R (2021) A TRIZ-based product-service design approach for developing innovative products. *Computers & Industrial Engineering* 161, 107608.
- Christensen BT and Schunn CD (2007) The relationship of analogical distance to analogical function and preinventive structure: The case of engineering design. *Memory & Cognition* 35, 29–38.
- Eilouti BH (2009) Design knowledge recycling using precedent-based analysis and synthesis models. *Design Studies* 30, 340–368.
- Friesike S, Flath CM, Wirth M and Thiesse F (2019) Creativity and productivity in product design for additive manufacturing: Mechanisms and platform outcomes of remixing. *Journal of Operations Management* 65, 735–752.
- Fu K, Chan J, Cagan J, Kotovsky K, Schunn C and Wood K (2013) The meaning of “near” and “far”: The impact of structuring design databases and the effect of distance of analogy on design output. *Journal of Mechanical Design* 135, 1–12.
- Fu K, Murphy J, Yang M, Otto K, Jensen D and Wood K (2015) Design-by-analogy: Experimental evaluation of a functional analogy search methodology for concept generation improvement. *Research in Engineering Design* 26, 77–95.
- Galle P (2009) The ontology of Gero’s FBS model of designing. *Design Studies* 30, 321–339.
- Giordano V, Puccetti G, Chiarello F, Pavanetto T and Fantoni G (2023) Unveiling the inventive process from patents by extracting problems, solutions and advantages with natural language processing. *Expert Systems with Applications* 229, 120499.
- Guo L, Yan F, Li T, Yang T and Lu Y (2022) An automatic method for constructing machining process knowledge base from knowledge graph. *Robotics and Computer-Integrated Manufacturing* 73, 102222.
- Han J, Shi F, Chen L and Childs PRN (2018a) The Combinator—a computer-based tool for creative idea generation based on a simulation approach. *Design Science* 4, e11.
- Han J, Shi F, Chen L and Childs PRN (2018b) A computational tool for creative idea generation based on analogical reasoning and ontology. *AI EDAM* 32, 462–477.
- Hao X, Ji Z, Li X, Yin L, Liu L, Sun M, Liu Q and Yang R (2021) Construction and application of a knowledge graph. *Remote Sensing (Basel Switzerland)* 13, 2511.
- Hao J, Zhao Q and Yan Y (2017) A function-based computational method for design concept evaluation. *Advanced Engineering Informatics* 32, 237–247.
- Haruna A, Yang M and Jiang P (2023) Design for additive manufacturing: A three layered conceptual framework for knowledge-based design. *Proceedings of the Institution of Mechanical Engineers. Part B, Journal of Engineering Manufacture* 237, 1405–1421.
- He B and Hua Y (2017) Feature-based integrated product model for low-carbon conceptual design. *Journal of Engineering Design* 28, 408–432.
- Hey J, Linsey J, Agogino AM and Wood KL (2008) Analogies and metaphors in creative design. *The International Journal of Engineering Education* 24, 283–294.
- Hu M, McComb C and Goucher-Lambert K (2023) Uncovering hidden patterns of design ideation using hidden Markov modeling and neuroimaging. *AI EDAM* 37, e8.
- Huang Z, Guo X, Liu Y, Zhao W and Zhang K (2023) A smart conflict resolution model using multi-layer knowledge graph for conceptual design. *Advanced Engineering Informatics* 55, 101887.
- Huang Z and Xie Z (2022) A patent keywords extraction method using TextRank model with prior public knowledge. *Complex & Intelligent Systems* 8, 1–12.
- Jia L, Peng Q, Tan R and Zhu X (2019) Analogical stimuli retrieval approach based on R-SBF ontology model. *Journal of Engineering Design* 30, 599–624.
- Jiang S, Hu J, Wood KL and Luo J (2022) Data-driven design-by-analogy: State-of-the-art and future directions. *Journal of Mechanical Design* 144(2), 1–55.
- Jiang S and Li J (2016) Research of the effectual action unit-based inverse method for solving the functional structure of design history. *Advances in Mechanical Engineering* 8, 1–12.
- Jiang S, Yang J, Xie J, Xu X, Dou Y and Jing L (2024) Product innovation design approach driven by implicit relationship completion via patent knowledge graph. *Advanced Engineering Informatics* 61, 102530.
- Jing L, Yang J, Ma J, Xie J, Li J and Jiang S (2023) An integrated implicit user preference mining approach for uncertain conceptual design decision-making: A pipeline inspection trolley design case study. *Knowledge-Based Systems* 270, 110524.
- Jing L, Zhang H, Dou Y, Feng D, Jia W and Jiang S (2024) Conceptual design decision-making considering multigranularity heterogeneous evaluation semantics with uncertain beliefs. *Expert Systems with Applications* 244, 122963.
- Li J, Guo X, Zhang K and Zhao W (2023) A knowledge-enabled approach for user experience-driven product improvement at the conceptual design stage. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 37, e22.
- Linsey JS, Wood KL and Markman AB (2008) Modality and representation in analogy. *AI EDAM* 22, 85–100.
- Liu L, Li Y, Xiong Y and Cavallucci D (2020a) A new function-based patent knowledge retrieval tool for conceptual design of innovative products. *Computers in Industry* 115, 103154–103154.
- Liu X, Luo L, Chen L and Li J (2020b) Research on bio-inspired design based on C-K theory for product innovation. *Journal of Mechanical Engineering* 56, 207–217.
- Liu W, Cai H, Cheng X, Xie S, Yu Y and dukehzyang (2022a) Learning high-order structural and attribute information by knowledge graph attention networks for enhancing knowledge graph embedding. *Knowledge-Based Systems* 250, 109002.
- Liu Q, Wang K, Li Y, Chen C and Li W (2022b) A novel function-structure concept network construction and analysis method for a smart product design system. *Advanced Engineering Informatics* 51, 101502.
- Liu H, Li Y, Bai Z and Wang Y (2023a) SMFM-based analogy retrieval tool for the conceptual design of innovative products. *Computers in Industry* 151, 103973.
- Liu S, Qin Y, Xu M and Kolmanič S (2023b) Knowledge graph completion with triple structure and text representation. *International Journal of Computational Intelligence Systems* 16, 1–12.
- Lu S, Guo Y, Huang W and Shen M (2022) Product form evolutionary design integrated with TRIZ contradiction matrix. *Mathematical Problems in Engineering* 2022, 1–12.
- Luo J, Sarica S and Wood KL (2021) Guiding data-driven design ideation by knowledge distance. *Knowledge-Based Systems* 218, 106873.
- Ma J, Hu J, Feng JF, Qi J and Peng YH (2016) Constrained FBS knowledge cell model, representation, and applications for conceptual design. *Proceedings of the Institution of Mechanical Engineers. Part C, Journal of Mechanical Engineering Science* 230, 1773–1786.
- Murphy J, Fu K, Otto K, Yang M, Jensen D and Wood K (2014) Function based design-by-analogy: A functional vector approach to analogical search. *Journal of Mechanical Design* 136(10), 101102.

- Nie Z, Cao G, Zhang P, Peng Q and Zhang Z (2022) Multi-analogy innovation design based on digital twin. *Machines (Basel)* **10**, 652.
- Peng G, Wang H, Zhang H, Zhao Y and Johnson AL (2017) A collaborative system for capturing and reusing in-context design knowledge with an integrated representation model. *Advanced Engineering Informatics* **33**, 314–329.
- Puccetti G, Chiarello F and Fantoni G (2021) A simple and fast method for named entity context extraction from patents. *Expert Systems with Applications* **184**, 115570.
- Rahman MH, Xie C and Sha Z (2021) Predicting sequential design decisions using the function-behavior-structure design process model and recurrent neural networks. *Journal of Mechanical Design* **143**(8), 081706.
- Romero Bejarano JC, Coudert T, Vareilles E, Geneste L, Aldanondo M and Abeille J (2014) Case-based reasoning and system design: An integrated approach based on ontology and preference modeling. *AI EDAM* **28**, 49–69.
- Sarica S, Luo J and Wood KL (2020) TechNet: Technology semantic network based on patent data. *Expert Systems with Applications* **142**, 112995.
- Shen T, Zhang F and Cheng J (2022) A comprehensive overview of knowledge graph completion. *Knowledge-Based Systems* **255**, 109597.
- Song H, Evans J and Fu K (2020) An exploration-based approach to computationally supported design-by-analogy using D3. *AI EDAM* **34**, 444–457.
- Song H and Fu K (2019) Design-by-analogy: Exploring for analogical inspiration with behavior, material, and component-based structural representation of patent databases. *Journal of Computing and Information Science in Engineering* **19**(2), 021014.
- Song H and Fu K (2022) Design-by-analogy: Effects of exploration-based approach on analogical retrievals and design outcomes. *Journal of Mechanical Design* **144**(6), 061401.
- Song B, Srinivasan V and Luo J (2017) Patent stimuli search and its influence on ideation outcomes. *Design Science* **3**, e25.
- Srinivasan V and Chakrabarti A (2011). Development of a catalogue of physical laws and effects using SAPPHIRE model. In *Design Creativity 2010* (pp. 123–130). Springer London.
- Srinivasan V, Song B, Luo J, Subburaj K, Elara MR, Blessing L and Wood K (2018) Does analogical distance affect performance of ideation? *Journal of Mechanical Design* **140**(7), 071101.
- Tan L and Zhang H (2021) An approach to user knowledge acquisition in product design. *Advanced Engineering Informatics* **50**, 101408.
- Valverde UY, Nadeau J-P and Scaravetti D (2017) A new method for extracting knowledge from patents to inspire designers during the problem-solving phase. *Journal of Engineering Design* **28**, 369–407.
- Vermaas PE and Dorst K (2007) On the conceptual framework of John Gero's FBS-model and the prescriptive aims of design methodology. *Design Studies* **28**, 133–157.
- Wang Z, Chen C-H, Zheng P, Li X and Khoo LP (2021) A graph-based context-aware requirement elicitation approach in smart product-service systems. *International Journal of Production Research* **59**, 635–651.
- Wang K, Tan R, Peng Q, Sun Y, Li H and Sun J (2021) Radical innovation of product design using an effect solving method. *Computers & Industrial Engineering* **151**, 106970.
- Wu C, Chen T, Li Z and Liu W (2021) A function-oriented optimising approach for smart product service systems at the conceptual design stage: A perspective from the digital twin framework. *Journal of Cleaner Production* **297**, 126597.
- Wu C, Zhou Y, Pereira Pessôa MV, Peng Q and Tan R (2021) Conceptual digital twin modeling based on an integrated five-dimensional framework and TRIZ function model. *Journal of Manufacturing Systems* **58**, 79–93.
- Zhang Y, Liu X, Jia J and Luo X (2020) Knowledge representation framework combining case-based reasoning with knowledge graphs for product design. *Computer-Aided Design and Applications* **17**, 763–782.
- Zhang P, Wang H, Li X, Nie Z and Ma Z (2023) Research on digital characterization and identification process model of functional genes for intelligent innovative design. *Advanced Engineering Informatics* **56**, 101983.
- Zhang Y, Wang H, Zhai X, Zhao Y and Guo J (2021) A C-RFBS model for the efficient construction and reuse of interpretable design knowledge records across knowledge networks. *Systems Science & Control Engineering* **9**, 497–513.
- Zheng H, Yang S, Lou S, Gao Y and Feng Y (2021) Knowledge-based integrated product design framework towards sustainable low-carbon manufacturing. *Advanced Engineering Informatics* **48**, 101258.
- Zhou B, Bao J, Li J, Lu Y, Liu T and Zhang Q (2021) A novel knowledge graph-based optimization approach for resource allocation in discrete manufacturing workshops. *Robotics and Computer-Integrated Manufacturing* **71**, 102160.
- Zhu G and Iglesias CA (2018) Exploiting semantic similarity for named entity disambiguation in knowledge graphs. *Expert Systems with Applications* **101**, 8–24.

Author biographies. Liting Jing received his Ph.D. degree in mechanical engineering from Zhejiang University of Technology in 2020. He joined Computer Science and Technology mobile station at Zhejiang University of Technology as a postdoctoral fellow in 2020. He joined the School of Mechanical Engineering at Zhejiang University of Technology in 2022. His research interests include product innovative design, design concept graphs, and conceptual scheme decision-making.

Mingyang Huang is a master's candidate at Zhejiang University of Technology. His research interests include product conceptual design, scheme optimization, and conceptual scheme decision-making.

Qizhi Li received his master's degree in mechanical engineering from Zhejiang University of Technology in 2023. He joined Xiān Lín Sānwēi Technology Company, focusing on product design modeling, performance optimization, and structural innovation.

Yubo Dou received his Ph.D. degree in mechanical engineering from Zhejiang University of Technology in 2024. He joined the Power Engineering and Engineering Thermophysics mobile station at Zhejiang University of Technology as a postdoctoral fellow in 2020.

Di Feng received his Ph.D. degree in mechanical engineering from Zhejiang University of Technology in 2022. He joined the School of Design and Architecture at Zhejiang University of Technology in 2022. His research interests include product innovative design, design concept graphs, and product service system design.

Shaofei Jiang is a professor at Zhejiang University of Technology. He received his Ph.D. in mechanical engineering from Zhejiang University in 2004. He has been awarded two second prizes for Scientific and Technological Progress in Zhejiang Province and one third prize for Scientific and Technological Progress from the China Machinery Industry Federation. His research interests include design theory and methodology, product conceptual design, product quality design, and mold design.

Appendix

A.1. Function basis classification

The introduction of function basis provides an effective tool for the expression of product functions, standardizes the expression of product functions, and prevents different designers from using different function words to describe the function of the same product. The first level of Fv is categorized into eight types: branch, guide, connect, control, transform, supply, detect, and support. The abstract level of Fv gradually decreases from the first level to the third level, as shown in Table A.1.

Table A.1. Three-level representation of Fv

First level	Second level	Third level
Branch	Separate	Sever, disconnect, off, refine, filter, purify, remove, and eliminate
	Distribute	Spill, splash, disperse, scatter, and divert
Guide	Input	Place, drop, and feed

(Continued)

Table A.1. (Continued)

First level	Second level	Third level
	Output	Expel, flow out, launch, eject, and extrude
	Convey	Transport and transmit
	Guide	Channel, derive, transform, revolve, rotate, flip, move, restrict, and obstruct
Connect	Link	Touch and hinge
	Blend	Stir and add
Control	Initiate	Start and activate
	Adjust	Regulate, control, and govern
	Enlarge	Increase, heat, accelerate, thicken, reinforce, strengthen, elevate, and pressurize
	Reduce	Decrease, down, thin, slenderize, and minimize
	Modify	Form, press, pierce, cut, crush, collide, throw, drill, squeeze, and pulverize
	Halt	Terminate, pause, interrupt, stop immediately, inhibit, shield, insulate, and protect
Transform	Switch	Liquefy, solidify, vaporize, and concentrate
Supply	Store	Receive, collect, gather, and absorb
	Furnish	Provide and replenish
Detect	Sense	Detect, identify, transduce, measure, and locate
	Indicate	Track, time, and schedule
Support	Stabilize	Secure, antislip, and fasten
	Ensure	Latch, clamp, and tighten

In addition, Fn is specifically divided into three levels. The first level of Fns is material, signal, and energy. The abstract level of Fn gradually decreases from the first level to the third level, as shown in Table A.2.

Table A.2. Three-level representation of Fn

First level	Second level	Third level
Material	Gas	Air, steam, nitrogen, methane, gas, etc.
	Liquid	Water, wastewater, pulp, solution, etc.
	Solid	Pipes, food, powder, debris, stones, wheels, etc.
	Mixture	Colloid, mixed gas, mixed liquid, mixed solid, solid-liquid mixture, etc.
Signal	State	Hearing, smell, touch, taste, sight, etc.
	Signal	DC current signal, AC current signal, pulse signal, voltage signal, power signal, etc.
	Acoustic	Ultrasonic, infrasound, tone, timbre, intensity, audio, etc.
	Light signal	Sunlight, light, wavelength, speed of light, etc.
Energy	Force	Pressure, shear force, torque, tension, bending force, gravity, suction, etc.
	Acoustic energy	Ultrasonic, infrasound, noise, white noise, etc.

(Continued)

Table A.2. (Continued)

First level	Second level	Third level
	Light energy	Ultraviolet, infrared, sunlight, solar energy, incandescent light, etc.
	Bioenergy	Charcoal powder, biogas, biodiesel and fuel ethanol.
	Electricity	Wind power, thermal power, direct current power, AC power, high frequency power, etc.
	Chemical energy	Battery power, oil, explosives, combustion, etc.
	Electromagnetic energy	Magnetic flux, magnetic field, electric field intensity, magnetic induction intensity, electric displacement, electric field energy density, etc.
	Mechanical energy	Kinetic energy, elastic potential energy, gravitational potential energy, rotation, translation, vibration, etc.
	Thermal energy	Temperature, heat, etc.

A.2. Dependency syntactic analysis

The LIP toolkit is specifically chosen because it is suitable for dependency syntactic analysis of Chinese text, which provides several functions such as lexical annotation, syntactic analysis, and semantic dependency analysis, and 17 DR tags have been defined in Table A.3.

Table A.3. Dependency syntax analysis grammar table

No.	DR tag	Explanation
1	SBV	Subject-verb
2	VOB	Direct object, verb-object
3	IOB	Indirect-object
4	FOB	Fronting-object
5	DBL	Double
6	ATT	Attribute
7	ADV	Adverbial
8	CMP	Complement
9	COO	Coordinate
10	POB	Preposition-object
11	LAD	Left adjunct
12	RAD	Right adjunct
13	IS	Independent structure
14	HED	Head
15	?DR	Optional DR
16	?/DR ₁ DR ₂ .../	Optional DR ₁ or optional DR ₂ ...
17	/DR ₁ DR ₂ .../	DR ₁ or DR ₂ ...

A.3. Pseudocode for domain distance calculation

The pseudocode accepts keywords and node types as inputs, establishes a connection to the Neo4j database, and leverages Cypher queries to retrieve node data and calculate domain distance. The pseudocode of node domain distance calculation based on Neo4j platform, as shown in Table A.4.

Table A.4. Pseudocode for calculating domain distance in Neo4j

Purpose: Calculate the domain distance between a given keyword and the specified node Input: Keyword, node type Output: Domain distance: D_p^t Parameter: $n-1$: node hops, o_k : node out-degree, i_k : node in-degree, D_o^t : absolute domain distance, D_{\max} : maximum domain distance.
1: Connect to the Neo4j database 2: Construct and execute a Cypher query to retrieve the $n-1$, o_k , and i_k of nodes related to the keyword Query result = nodes_info 3: Initialize D_o^t and D_{\max} to 0 4: For each node retrieved: a: Calculate node's degree sum = $o_k + i_k$ b: Calculate D_o^t based on Eq. (5) with node's ID, $n-1$, and degree sum 5: Update D_{\max} if $D_o^t > D_{\max}$ 6: After all nodes are processed, calculate D_p^t based on Eq. (6) with D_o^t and D_{\max} 7: Disconnect from Neo4j database 8: Return domain distance D_p^t

A.4. Classification and cluster analysis of Fv

Taking PIR as an example, Fv in the technical background of patent text are extracted as data sources. According to Eq. (2) weight calculation formula, the top 50 verbs with frequent occurrence are listed in detail as technical keywords, and the relative importance of each verb in technical description is quantified. The word frequency weights of patent technical background verbs are shown in Table A.5.

Use Word2Vec to vectorize the top 50 technical keywords of PIR invention patents. We performed clustering analysis based on cosine similarity, and the results are shown in Figure A.1 (translated from Chinese into English).

Table A.5. Top 50 technical verbs and their weights in calculation results

Verb (V_i)	Word frequency (c_i)	Relative weight (w_i)	Verb (V_i)	Word frequency (c_i)	Relative weight (w_i)	Verb (V_i)	Word frequency (c_i)	Relative weight (w_i)
Detect	1698	0.1750	Plug	164	0.0169	Ruin	90	0.0093
Drive	736	0.0759	Repair	163	0.0168	Transmit	90	0.0093
Walk	559	0.0576	Extend	162	0.0167	Affected	88	0.0091
Control	400	0.0412	Creep	162	0.0167	Impair	86	0.0089
Convey	331	0.0341	Explode	158	0.0163	Examine	85	0.0088
Maintain	299	0.0308	Service	147	0.0152	Disclose	82	0.0085
Operate	272	0.0280	Monitor	136	0.0140	Turn	78	0.0080
Transport	250	0.0258	Regulate	131	0.0135	Weld	73	0.0075
Leak	243	0.0250	Recognize	122	0.0126	Pull	69	0.0071
Corrode	240	0.0247	Measure	116	0.0120	Replace	61	0.0063
Support	233	0.0240	March	113	0.0116	Seep	52	0.0054
Collect	230	0.0237	Worm	113	0.0116	Crack	46	0.0047
Clean	214	0.0221	Unblock	112	0.0115	Push	46	0.0047
Carry	202	0.0208	Fix	108	0.0111	Drain	46	0.0047
Block	171	0.0176	Remote control	104	0.0107	Slip	45	0.0046
Damage	170	0.0175	Obstacle cross	99	0.0102	Bend	44	0.0045
Inspect	170	0.0175	Break	92	0.0095			

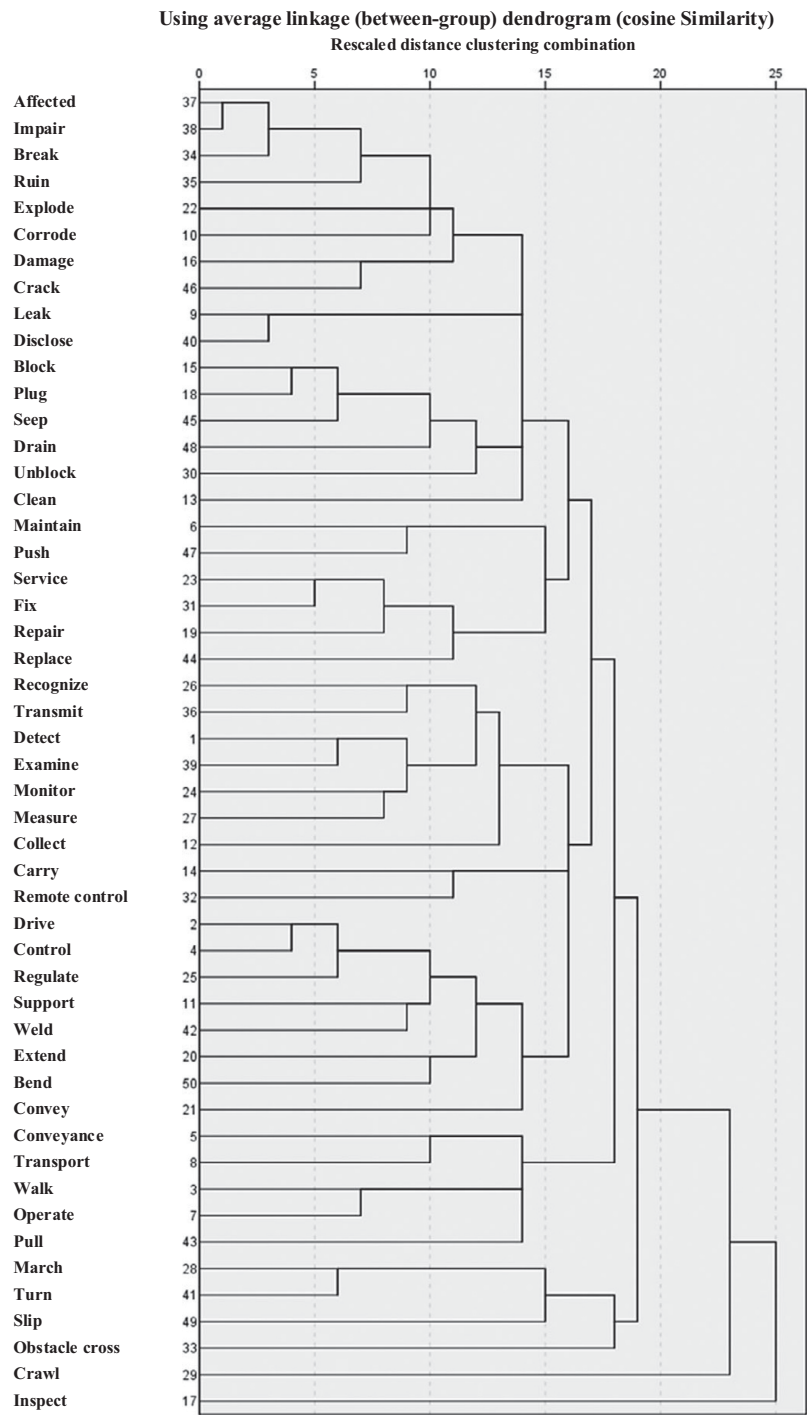


Figure A.1. Keyword cluster analysis based on cosine similarity.