Extraction and linking of motivation, specification and structure of inventions for early design use

Pingfei Jiang, Mark Atherton & Salvatore Sorce

Published online: 23 Jun 2023.

Submit your article to this journal

Article views: 87

View related articles

View Crossmark data

Taylor & Francis
Taylor & Francis Group

Extraction and linking of motivation, specification and structure of inventions for early design use

Pingfei Jiang ⬤[a], Mark Atherton[b] and Salvatore Sorce ⬤[c]

[a]Department of Mechanical Engineering, School of Engineering and the Environment, Kingston University London, London, UK; [b]Department of Mechanical and Aerospace Engineering, College of Engineering, Brunel University London, London, UK; [c]Faulty of Engineering and Architecture, Koew University of Enna, Enna, Italy

ABSTRACT

Novel design creations often exist in the form of patents. It is well acknowledged that patents are a great source of design inspiration, therefore designers are encouraged to engage patents early in the design process. Studies of patent analysis have been carried out to benefit engineering design activities such as patent classification, technology forecasting, idea generation and emerging design-prior art comparison. However, the design intent behind inventions has received little attention in patent analysis. Designers can gain better design insight by looking at a patent from a systematic perspective starting from design intent to principal solutions. In this paper, an approach is proposed to extract and link the knowledge conveyed within patent descriptive sections to typical early engineering design stages, namely Motivation, Specification and Structure. This knowledge is then conceptualised into TRIZ engineering parameters and reconciled functional basis to enable cross-patent analysis. When compared to expert analysis, the proposed approach achieves an average of 63% accuracy with respect to Motivation, 56% with respect to Specification function and 44% with respect to Specification flow. Potential applications of such linking for early design use are then demonstrated through a prototype knowledge network.

1. Introduction

Innovation plays a crucial role in achieving the economic growth and development of technology-based firms (Han et al. 2021). Design innovation often results in patents, offering public disclosure in exchange for legal coverage (Roca-Gonzalez, Vera-Lopez, and Rodriguez-Bermudez 2018). Meanwhile, studying patents also have a significant influence on design innovation (Koh 2020; Siddharth, Li, and Luo 2022c). As a large and freely accessible information source, patents offer a promising potential source of design inspiration to support enhanced ideation (Luo, Sarica, and Wood 2019; Wodehouse et al. 2017; Wodehouse et al. 2018). More than 90% of inventions across the world can be found in patents,

CONTACT Pingfei Jiang ✉ p.jiang@kingston.ac.uk ⊞ Department of Mechanical Engineering, School of Engineering and the Environment, Kingston University London, London SW15 3DW, UK

with 80% of the techniques revealed by these patents not found in other professional articles (Chen 2009). A range of activities throughout the product development process can be supported by patents, for example, scoping, generation and embodiment (Jiang and Luo 2022; Vasantha et al. 2017). Valuable information on engineering design concepts and their physical realisations can also be obtained through the study of patents (Li et al. 2012).

For a designer to obtain an insightful understanding of patents, it is beneficial that the invention design knowledge can be represented in a systematic manner that suits the way of the designer's thinking. Ríos-Zapata et al. (2017) summarised 6 phases of product development: need identification, product requirements, conceptual design, embodiment design, detailed design and manufacturing design, and classified the first three phases as early stages which are used in this study. 'Need identification' provides valuable insight into what is the driver for the design (Howard, Culley, and Dekoninck 2008). This can be reflected in customer desires, government policies and customer reviews (Dieter and Schmidt 2012). In the context of patent analysis, it is beneficial to know why an invention exists and what problems it addresses to enable an improved understanding. The main goal of 'Product requirements' is to identify a series of functions that the product must accomplish (Ríos-Zapata et al. 2017). 'Conceptual design' determines the principal solutions to fulfil the series of functions. At this stage quantity of ideas is desired over quality to increase the likelihood of achieving a desirable product (Liu et al. 2020). The study and analysis of patents can be time-consuming and a challenge for designers without the aid of data analysis tools. For example, the rich technical content within a patent can be hidden within the legal terminologies used and thus require considerable effort and expertise to extract (Kim, Suh, and Park 2008). Moreover, as of 2018, the number of worldwide patent applications has been growing yearly by 8.3% (Sorce et al. 2018), which highlights a further increase in the availability of patent knowledge.

Over recent decades, due to general advances in artificial intelligence and data science, there has been a rapid growth in data science research specifically applied to engineering design. A recent review study carried out by Chiarello, Belingheri, and Fantoni (2021) shows that only around 35% of engineering design publications contain data science topics, amongst which only around 15% of the publications relate to the problem definition ('Need identification' and 'Product requirements') stage of engineering design. On the contrary, another review study carried out by Jiang et al. (2022b) shows that more than 70% of patent-for-design publications contribute to design methodology, i.e. the development of a method or framework that can be applied to assist design activities, which would be expected to cover the early stages of design. This indicates a lack in the application of data science in patent-for-design research for early design use, especially in the problem definition stage. Therefore, in this paper the authors develop an approach to identify and map invention design knowledge to three early design stages, need identification, product requirements, conceptual design, with a particular focus on the first two stages, aiming to provide designers with richer insight into existing inventions. Three types of knowledge will be referred to as **Motivation** (relates to Need identification), **Specification** (relates to Product requirements) and **Structure** (relates to Conceptual design).

2. Related work

This study builds on prior studies of patent analysis for engineering design and natural language processing of patent documents to link the motivation, specification and structure of inventions. Relevant literature on these topics is presented in this section.

2.1. Patent analysis for engineering design

An early review carried out by Abbas, Zhang, and Khan (2014) summarised applications of patent analysis. Since then, more recent literature has demonstrated that studying patents can benefit the development of products in various aspects, for example, technological forecasting (Cho et al. 2021 Han et al. 2021; Higham, Contisciani, and De Bacco 2022; Yang et al. 2022) and patent classification (Grawe, Martins, and Bonfante 2017; Henriques, Ferreira, and Castelli 2022; Hu et al. 2018; Jiang et al. 2022a; Yan and Luo 2017). A survey on deep learning for patent analysis carried out by Krestel et al. (2021) classifies the most popular tasks for automated patent analysis, however, design is not explicitly addressed.

From the perspective of engineering design, research has mostly concentrated on using patents to assist designers in performing design-centred tasks. Jiang et al. (2022b) categorise four applications of patent analysis in engineering design research: design theory, design methodology, design tool and design strategy. The study shows that design methodology is the most popular theme of research contribution. Design methodology can be understood as methods or frameworks that can be repeatably applied to aid design activities. A popular stream of research is to use appropriate forms of representation to facilitate a designer's understanding of patents that can lead to design insight. Yang and Soo (2012) use Part-Of-Speech and dependency trees to construct concept graphs from patent claims. Fantoni et al. (2013) extract design information from patents using the function-behaviour-state (FBS) model to provide additional graphical visualisations alongside patent drawings. Yuan et al. (2016) use qualitative processing reasoning to decompose an overall function of design to obtain the function structure and principle solution using SysML. Vasantha et al. (2017) also utilise the FBS model to represent patent abstracts and use crowdsourcing as a means to evaluate people's understanding of patents. Jiang et al. (2017) further develop the functional analysis diagram (Aurisicchio, Bracewell, and Armstrong 2013) to represent patent working principles and later automate the whole process (Jiang, Atherton, and Sorce 2021) to produce design insight.

Studies also focus on cross-patent knowledge representation. Kang et al. (2015) propose a methodology of using a patent-function matrix to reveal the distribution of patents across technical functions. Atherton et al. (2017) extract functional interactions between geometric features in patents to create graphical and semantic annotations of patent drawings. Luo et al. (2018) use patent IPC to develop a cloud-based system, InnoGPS, to enable designers to navigate technical concepts revealed in patents. Sarica, Luo, and Wood (2020) develop TechNet, a semantic network of engineering concepts based on the analysis of patent title and abstract, to reveal their semantic associations. Siddharth et al. (2022b) develop an approach to obtain an engineering knowledge graph containing both common sense and engineering inferences.

Some research focuses on design for patentability, i.e. avoiding potential conflicts with existing IP, for instance, TRIZ-led patent mapping to identify potential conflicts between

patents (Li, Atherton, and Harrison 2014), a TRIZ-based patent trimming design around method (Li et al. 2015), a re-design framework based on function analysis (Cheng et al. 2016), a framework for emerging design-prior art comparison enabled by functional geometry interaction and domain-specific ontology (Jiang et al. 2018), a design around approach using both evolution theory and bundle-type patent portfolio analysis (Li et al. 2019a) and a re-design process based on reversing engineering and patent circumvention (Akerdad, Aboutajeddine, and Elmajdoubi 2022).

Patent analysis for engineering design is also broadly applied to aid ideation and hence promote innovative designs. Murphy et al. (2014) develop patent-based functional analogy search to aid concept generation. This approach is evaluated, showing significantly improved solutions in novelty (Fu et al. 2015). A research express solution finder (RESF) is proposed by Ríos-Zapata et al. (2017) to search for adequate solutions for subsystem designs through patent analysis. Valverde, Nadeau, and Scaravetti (2017) develop a three-stage methodology using pertinent keywords as the basis of knowledge for patent retrieval and use a discovery matrix to identify design opportunities. Sorce et al. (2018) develop a visual interface inspired by block-oriented programming to promote design innovation which is subsequently shown to be effective (Sorce et al. 2019). Song, Luo, and Wood (2019) propose a method that generates and uses a network of functions through their co-occurrences in patents to identify a core–periphery structure. A study carried out by Liu et al. (2020) presents cross-domain patents that can be retrieved using matching functional basis keywords from the abstract for the conceptual design of innovative products. Chan et al. (2021) use functional breakdown and a TRIZ scientific effect database to retrieve similar patents using Doc2Vec model, such that more updated and specific examples can be provided to support the generation of innovative solutions. Liu, Li, and Li (2021) transform IPC text into effects described by a hierarchical function verb taxonomy, enabling designers to map functional requirements to desired effect knowledge. Sarica et al. (2021) use TechNet to stimulate idea generation by encouraging designers to infer and relate new concepts beyond the domain of interest, hence promoting the generation of new ideas. Luo, Sarica, and Wood (2021) propose a knowledge-based system that provides design stimuli according to knowledge distances across engineering fields.

To date, literature has provided valuable knowledge regarding how patent analysis can benefit engineering design by providing design insight and supporting ideation. However, the majority of research utilises functional attributes of designs either as input or output of their analysis, meaning that the purpose of an invention is overlooked. The authors believe that when engaging with patents, it is beneficial to know the reason why a particular invention is disclosed, in other words, what the motivation is behind it and how it is addressed by the design. Jack (2013) suggests that the background section of a patent can be used to identify needs and design motivation. In the work of Cascini and Zini (2008), the problem statement is mentioned. However, these problem statements are processed manually by expert analysis, which possess challenges when analysing a large number of patents. 'Needs' were mentioned by Fantoni et al. (2013) and referred to as the voice of customers but were not addressed in the approach developed. By incorporating the analysis of invention motivation, patent knowledge can be extracted and represented that fits the natural way of designers' thinking. Furthermore, prior studies that focus on cross-patent analysis enabled by networks do not provide traceability of patents, meaning designers cannot identify the source patent for further examination.

2.2. Natural language processing of patents

Natural language processing (NLP) schemes have been developed in the past years, and aim to contribute to design research (Siddharth, Blessing, and Luo 2022a). NLP has been broadly applied in patent analysis to extract data from patents that are related to design. For example, Li et al. (2012) used NLP techniques to estimate the TRIZ level of invention for better classification. Fantoni et al. (2013) apply NLP to extract function, behaviour and state information from patent texts, enabling a graphical visualisation of a patent. Cao et al. (2016) use NLP to extract technical system components and their relationships to construct a design structure matrix. Li and Tate (2019) use part-of-speech tagging and statistical parsing techniques to identify functional requirements and design parameters. Jiang, Atherton, and Sorce (2021) utilise part-of-speech tagging and regular expression parsing to achieve automated patent functional modelling, by identifying Subject-Action-Object triplets from the patent-independent claim. In this section, popular techniques of NLP, sentiment analysis and word embeddings are reviewed.

Sentiment analysis, also known as opinion mining, is a method often used in NLP that identifies the degree of negativity of sentences. Sentiment analysis is broadly applied in analysing people's comments on social media (Li et al. 2019b; Soong et al. 2019) and product/service reviews (Han and Moghaddam 2021; Jeong, Yoon, and Lee 2019; Zvarevashe and Olugbara 2018), to help organisations to make strategic decisions. As far as the authors are aware, sentiment analysis has not been applied to patent analysis. Sentences in patent descriptive sections that carry negative sentiment should possess a reasonable chance of describing shortcomings of the prior art. Identifying shortcomings of prior art implies that the present invention disclosed should be able to tackle those shortcomings, hence the motivation for the invention.

Word embedding models vectorise textual phrases to capture useful semantic properties and linguistic relationships between words (Wang et al. 2018). Word embeddings can enable a broad range of generic NLP tasks such as text generation (Qu et al. 2020), question answering (Esposito et al. 2020) and text classification (Stein, Jaques, and Valiati 2019). Word embeddings have also been applied to patent classification, for example, Grawe, Martins, and Bonfante (2017) present a patent classification approach using word embedding and Long Short Term Memory. Hu et al. (2018) develop a patent keyword extraction algorithm based on the distributed Skip-gram model for improved patent classification. Jiang et al. (2022a) combine word embeddings of patent title and abstract together with patent image feature extraction using conventional neural networks to achieve a patent classification architecture named TechDoc. Henriques, Ferreira, and Castelli (2022) apply machine learning with transfer learning on second-level IPC to classify Portuguese patents. Studies also focus on using word embeddings for named entity recognition. Habibi et al. (2017) apply deep learning and word embeddings in biomedical named entity recognition (NER). Thorne and Akhondi (2020) compare the qualities of word embeddings for NER by using training patents from different domains. Word embeddings are also broadly applied in the construction of patent network graphs, for instance, TechNet (Sarica, Luo, and Wood 2020), Engineering Knowledge Graph (Siddharth et al. 2022b), WikiLink (Zuo et al. 2022) and Design Knowledge Graph (Sarica, Han, and Luo 2023).

Despite various applications of word embeddings in patent analysis, little research has focused on how word embeddings can be used to establish semantic associations within

and across patents that involve invention motivation. This study aims to address afore-mentioned gaps with a particular focus on identifying and linking invention motivation, specification and structure, to provide novel insight at early stages of design. In the next section development of the approach is presented.

3. Development of the approach

3.1. Data collection

The first step is to determine where the patent data should be collected from, i.e. which patent search engine to use. The most used English patent search engines include Google Patents, EPO and USPTO. Both Google and EPO have access to global patents whilst USPTO tends to only cover patents filed in the US region. Compared to EPO, Google patents is quicker and more familiar to use, and can load patents in full text (Marley 2015) hence is used in this study.

To get started, initial searches using Google Patents are required to obtain a list of patents to be analysed. Then the results are exported in the form of a spreadsheet using the 'Download' feature in Google Patents, which contains information such as title, date, assignee and links of each patent. The spreadsheet is then imported into Python, allowing web scraping of patent contents using the links. Python module BeautifulSoup with html parser is used to carry out the web scraping. The scraped contents are stored in Python, ready for the next step.

3.2. Extraction of motivation specification and structure

Although patent background sections are most suited for motivation recognition, it has limited information on the full patent. Moreover, from a web scraping viewpoint, it is difficult to extract the background section solely as patents use different html classes on Google Patents, making them difficult to scrape. Thus, the entire description section is used to recognise motivation. Firstly, sentences that contain clear motivation indicators such as 'there is a need for … ' or 'it is desired that … ' are extracted using string matching. Then, sentiment analysis is carried out for each sentence of the patent description to identify those carrying negative sentiment. A pre-trained sentiment analysis model developed by Hugging Face, 'distilbert-base-uncased-finetuned-sst-2-english' is used to predict the sentiment scores, in the form of a label and a score. Sentences labelled 'negative' and with a score that is higher than 0.999 are considered to be motivation sentences. These sentences are then joined with the sentences containing motivation indicators and then compiled into motivation sentences. Figure 1 shows the procedure for extracting motivation from a patent.

The patent-independent claim is used to extract invention Specification and Structure. This is because, by law, the independent claim must be self-contained and fully defines the scope of the invention. IPO (2020) states that the distinctive technical features must be set out by the first independent claim, distinguishing the invention from the prior art. Therefore, analysing the independent claim should provide sufficient information to help the designer obtain insight and decide whether to investigate a patent in detail, rather than targeted to capture the full patent. Invention Structure, identified in the form of technical

Extract motivation sentences
Input:
> *Retrieve scrapped patent description content*

Procedure:
> *Sentence tokenisation*
> *Start loop:*
>> *FOR each sentence token:*
>>> *Implement sentiment analyser*
>>> *IF Sentiment Label = "negative" AND Sentiment Score > 0.999*
>>>> *Add to Motivation*
> *End loop*

Output:

Figure 1. Procedure of extracting motivation from a patent.

systems/components revealed by the abstract, is expected to be expressed in compound noun phrases, e.g. 'diesel engine', 'set of batteries' and 'at least one power supply'. By law, a patent-independent claim can only have one sentence, therefore, in this study, invention Specification is presumed to be expressed as short sentences separated by either a comma or semicolon, and within these short sentences, at least one action (verb) should be performed. Examples of extracted Specification include 'a nozzle for a fan assembly for creating an air current,' and 'a collapsible front tube assembly having a first end and a second end supporting a front wheel.'

Extraction of Specification starts with customised sentence tokenisation by using Natural Language Toolkit (NLTK). End characters '.', ':', ',' and ';' are used to split short sentences from the independent claim. Then the stop words from each short sentence are removed by referring to a customised list (see Table 1). The use of a customised list is because stop words that are built in the NLTK corpus contain too many meaningful preposition phrases such as 'in', 'of', 'from', and 'to', making it inaccurate for analysis if used.

Then a Part-Of-Speech (POS) tagger is built using scikit-learn in Python. The Penn Treebank Corpus from NLTK is used to train the POS tagger. 80% of the dataset is used as training sentences and 20% is used for testing. Features of the tokenised word including the previous word, the next word, 1 to 3-letter prefix and suffix of the word are taken into consideration during training. DecisionTreeClassifier from scikit-learn is used with 20k samples. The trained POS tagger against the testing dataset achieved an accuracy of 90.8%. This tagger is then used to carry out noun phrase lemmatisation, to consolidate noun phrases with similar expressions such as 'battery' and 'batteries'. This is accomplished by converting POS tags to WordNet tags and then applying the WordNet Lemmatiser. The reason for lemmatising noun phrases only while keeping other phrases unchanged is to maintain the accuracy of parsing performed later. A classifier-based chunker is then built using the CoNLL 2000 corpus. 85% of the dataset is used for training and 15% used for testing. The trained chunker achieved a 93.1% accuracy in identifying IOB (Inside, Outside, Beginning) tags and an F-measure of 89.2%. The trained chunker works well in identifying nouns and verbs with

Table 1. List of customised stop words used for independent claims.

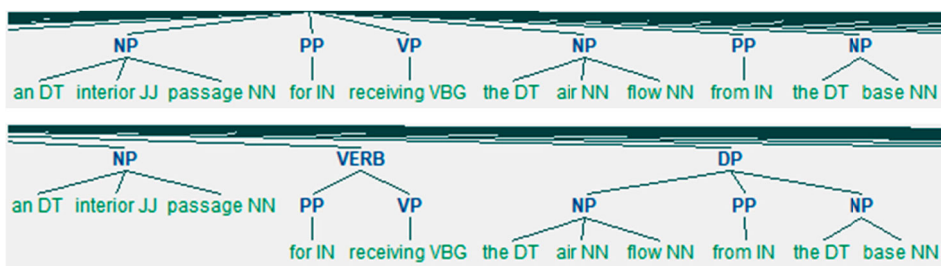| said	this	that	these	those	which	what

Figure 2. An example comparison using classifier-based chunker only (top) and plus an addition step of shallow parsing using regular expression (bottom) for patent US9249810B2.

Table 2. List of hierarchy words used for specification.

comprising	consist	consists	comprise	comprises	consisting	having
has	have	Includes	including	include	consists of	composed of
consist of	composing	Composes	containing	contain	contains	

labels 'NP' and 'VP' respectively, however, Structures that are expressed in a more complex form might be missed as well as Specification that includes prepositions, e.g. 'to define' and 'created by'. As a result, an additional step of shallow parsing using regular expression is performed to capture the more intricate form of Structure and Specification. For instance, 'DP', referred to as design parameter, can exist in the form of $<NP><PP><NP><NP>*$, "VERB", referred to as verb, can exist in the form of $<PP>?<VP>++<RB>*<PP>*$. This results in another tree consisting of larger chunks. Figure 2 shows an example comparison of chunks using only the classifier-based chunker (top) and with an additional step of shallow parsing (bottom) for a bladeless fan patent independent claim US9249810B2. 'the air flow', 'from' and 'the base' is identified as one design parameter 'the air flow from the base', 'for' and 'receiving' is identified as one verb 'for receiving'.

So far, each tokenised sentence is converted into a chunked tree with corresponding labels. The next step to identify Specification is to check whether each tree has a VERB label in it. If so, it will be stored for further processing. Leaves with the label 'NP' or 'DP' are considered as Structure, e.g. 'an interior passage', 'the air flow from the base' in Figure 2. Each stored sentence token will then be checked against a customised hierarchy term list (see Table 2) to see whether that sentence is only describing a hierarchical relationship. If a sentence token only has one VERB and that VERB is within the list of hierarchy words, that sentence token is then removed from the specification list. Figure 3 shows the procedure of identifying Specification and Structure from patent independent claim.

3.3. Linking motivation, specification and structure

So far invention Motivation, Specification and Structure have been extracted using patent description and independent claim, expressed in natural language. In order to facilitate cross-patent analysis and discovery, Motivation and Specification need to be associated with standardised libraries of expressions. In this study, the TRIZ 39 engineering parameters (Mann 2001) and reconciled functional basis (RFB) (Hirtz et al. 2002) are used. The

```
## Extract specification sentences ##
Input:
        Retrieve scrapped patent independent claim

Procedure:
        Customised sentence tokenisation
        Customised stop words removal
        POS tagger training
        Chunker training
        Define further parsing grammar
        Start loop:
                FOR each sentence token:
                        POS tagging
                        Two-stage chunking
                        IF chunked tree has one VERB label AND VERB is a hierarchy term
                                Remove from Specification
                        ELSE
                                Add to Specification
        End loop
Output:
        List of Specification


## Extract structure terms ##
Input:
        Retrieve scrapped patent independent claim

Procedure:
        Customised sentence tokenisation
        Customised stop words removal
        POS tagger training
        Chunker training
        Define further parsing grammar
        Start loop:
                FOR each sentence token:
                        POS tagging
                        Chunking
                        Two-stage chunking
                        IF subtree label = DP or NP:
                                Add to Structure
        End loop
Output:
        List of Structure.
```

Figure 3. Procedure of extracting specification and structure from a patent

TRIZ 39 engineering parameters refer to improving and worsening features of a system to guide the generation of ideas, which can indicate the improvement claimed by an invention compared to the prior art. RFB is a standardised set of terms to represent product technical functions in engineering design. It uses the concept of function and flow to provide consistency in communication between designers. RFB uses a three-tier system (primary, secondary and tertiary) to incorporate both function and flow terms. In this study, secondary terms plus the power conjugate complements for energy terms are used for the convenience of data analysis. Figure 4 presents the linking approach to link invention Motivation, Specification and Structure. The linking approach is accomplished in two stages, conceptualisation and mapping.
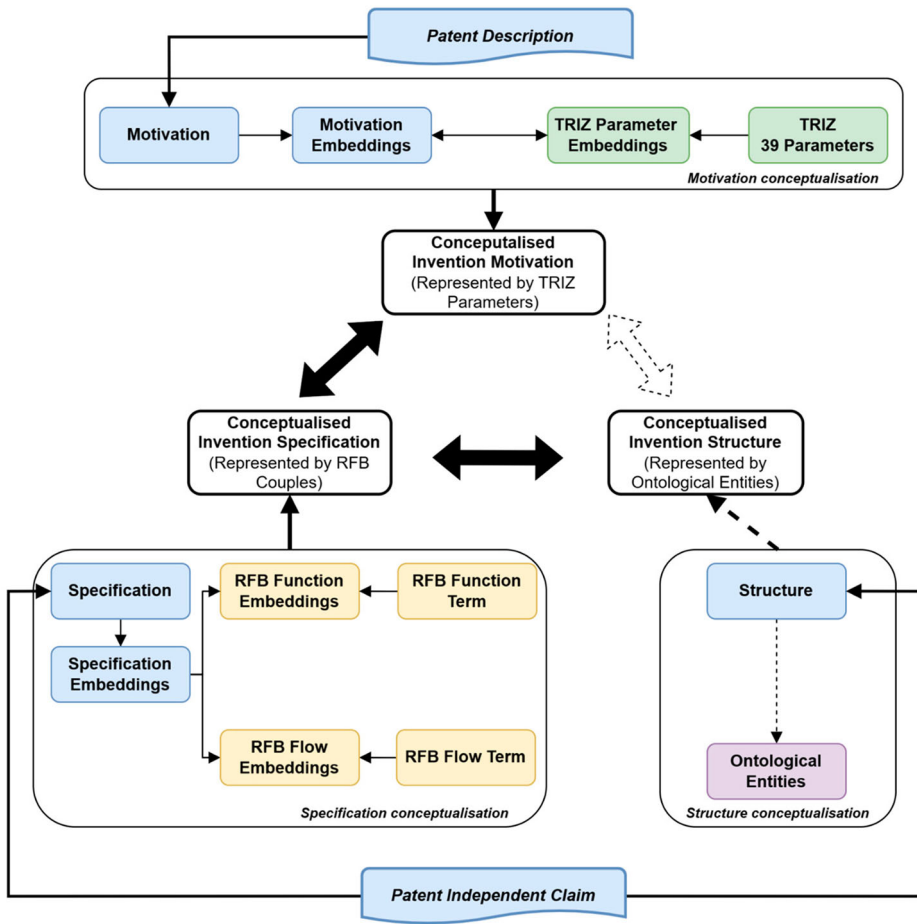
Figure 4. Linking approaches for linking invention motivation, specification and structure.

3.3.1. Conceptualisation

Extracted motivation from the previous step is vectorised using Sentence-BERT (Reimers and Gurevych 2019), a well-known Python framework for sentence embeddings. A pre-trained model 'all-mpnet-base-v2', available on Hugging Face, with an average performance of 63.3% on both sentence embeddings and semnac search, is used to convert each motivation sentence to a 768-dimensional dense vector. For the 39 TRIZ parameters, with the aim of better defining what each parameter means, their corresponding explanation (Gadd 2011) is used instead of the parameter names. For example, instead of using 'Speed', its explanation, 'The velocity of an object; the rate of a process or action in time' is used. After both motivation and TRIZ parameters are vectorised, Pearson's correlation coefficient is calculated for each motivation embeddings against all 39 TRIZ parameter embeddings. The TRIZ parameter whose embeddings yields the highest Pearson's coefficients is considered to be the conceptualisation of that particular motivation, see Figure 5 for this conceptualisation procedure.

A similar approach is applied for the conceptualisation of Specification. First of all, each extracted Specification sentence token is vectorised using Sentence-BERT. For each RFB

Conceptualisation of Motivation
Input:
    Extracted motivation sentences
    TRIZ 39 parameters
    TRIZ 39 parameter explanations
Procedure:
    Fetch pre-trained sentence embedding model
    Create embedding for all motivation sentences
    Create embedding for TRIZ 39 parameter explanations
    Start loop:
        FOR each motivation sentence embedding:
            Compute Pearson's ratio (motivation, TRIZ embedding FOR all TRIZ parameters)
            Rank results to obtain MAX (Peason's ratio) and its corresponding TRIZ parameter
    End loop
Output:
    List of Motivation and their correspond TRIZ parameter

Figure 5. Procedure of motivation conceptualisation using TRIZ parameter and their explanations.

secondary function and flow terms, their corresponding explanation (Hirtz et al. 2002) was used to obtain their embeddings. For example, instead of using function 'actuate', its explanation, 'To commence the flow of energy, signal, or material in response to an imported control signal' is used. However, it was found that RFB flow terms are already self-explanatory compared to their explanation, for example, 'force' is already quite clear compared to its explanation 'The action that produces or attempts to produce a translation'. Therefore, it was decided that RFB flow terms are used rather than their explanations. After specification, RFB function and RFB flow are vectorised, Pearson's correlation coefficient is used again to measure the similarity between each specification embedding against each RFB function and flow embedding. This results in one RFB function term and one RFB flow term whose embeddings yield the highest Pearson's ratio. The couple formed by these function and flow terms is ultimately considered to be the conceptualisation of that particular specification. Figure 6 shows the procedure of specification conceptualisation.

Conceptualisation of the invention Structure extracted earlier can be achieved with the aid of knowledge representation schemes such as WordNet (Miller 1995), ConceptNet (Speer, Chin, and Havasi 2017) and TechNet (Sarica, Luo, and Wood 2020), amongst which TechNet appears to be more suitable for engineering design. However, implementation of such schemes requires a considerable amount of time and resources hence is not addressed in this paper, as indicated by a broken arrow in Figure 4. As a result, at the moment invention Structure is expressed in patent terms.

3.3.2. Mapping

Mapping between motivation and specification is achieved by computing Pearson's correlation coefficient between extracted motivation and specification sentences. This will more accurately reflect patent-specific knowledge association. However, with the aim of supporting cross-patent analysis and discovery, the established association will be passed onto their corresponding conceptualised terms. Mapping between specification and structure is based on the chunked tree obtained earlier. Subtree with the label 'VERB' will be used as the centre for Structure identification following a Subject-Action-Object format in which

Conceptualisation of Specification
Input:
>> *Extracted motivation sentences*
>> *RFB function and flow terms*
>> *RFB function and flow term explanations*

Procedure:
>> *Fetch pre-trained sentence embedding model*
>> *Create embedding for all specification sentences*
>> *Create embedding for RFB function using their explanations*
>> *Create embeddings for RFB flow*
>> *Start loop:*
>>> *FOR each specification sentence embedding:*
>>>> *Compute Pearson's ratio_Function (specification, RFB function embeddings)*
>>>> *Compute Pearson's ratio_Flow (specification, RFB flow embeddings)*
>>>> *Rank results to obtain MAX (Peason's ratio_Function) and its corresponding RFB function terms*
>>>> *Rank results to obtain MAX (Peason's ratio_Flow) and its corresponding RFB flow terms*
>> *End loop*

Output:
>> *List of specification and their correspond RFB function and flow terms*

Figure 6. Procedure of specification conceptualisation using RFB function and flow.

'VERB' is the 'Action'. The algorithm will navigate through each tree, and locate every 'VERB' and the 'NP' or 'DP' on its near left and right. By doing so the mapping can be achieved.

With each patent's Motivation, Specification and Structure conceptualised and linked, patents can be analysed in bulk to provide further insight into the early design stages. For example, the analysed results can enable a knowledge base of interlinked invention Motivation, Specification and Structure for designers to explore. Figure 7 shows the procedure of how the mapping is accomplished and a knowledge network can be constructed. The association between each knowledge element, i.e. node, is simply the data pairs obtained through the mapping using Pearson's correlation hence is not weighted.

4. Method evaluation

In this section, the proposed method is applied to 15 patents from five engineering design invention categories (see Table 3). Only three patents per category because of the time required to perform thorough analysis for each patent. When performing HTML scrapping, for consistency purposes, only granted patents from the US patent office were selected. The invention categories specified in Table 3 were used as Google patents search keywords, e.g. 'bladeless fan', and 'hybrid locomotive'. Then three patents were selected randomly from the results and their corresponding weblink is stored for later access. Notation for each patent entry is assigned.

Extraction and conceptualisation results obtained from these 15 patents were evaluated against expert analysis as well as zero-shot text classification (Yin, Hay, and Roth 2019). The expert analysis was performed by two designers together, one with over 25 years of experience and another one with over 8 years of experience. The results shown in the section are the agreed outcomes. Worth noticing that in order to establish a standard procedure

Mapping between Motivation and Specification
Input:
     Extracted motivation sentences and their TRIZ representation
     Extracted specification and their RFB representation
Procedure:
     Start loop:
         FOR each motivation sentence:
            Fetch its sentence embeddings and TRIZ representation
            Compute Pearson's ratio (motivation, specification of ALL specifications)
            Rank results to obtain MAX (Pearson's ratio) and its corresponding specification
            Fetch the RFB representation for the selected specification
     End loop
Output:
     Data pairs of (motivation, specification), (TRIZ, RFB)

Mapping between Specification and Structure
Input:
     Chunked tree list for each patent independent claim
     Extracted specification and their RFB representation
Procedure:
     Start loop:
         FOR subtree in Chunked tree:
            Fetch its corresponding specification and RFB representation
            Locate VERB in subtree
            Identify the next NP/DP on the left of VERB
            Identify the next NP/DP on the right of VERB
     End loop
Output:
     Data pairs of (specification, SAO triples)

Construct knowledge network
Input:
     Data pairs of (motivation, TRIZ), (specification RFB), (motivation, specification), (TRIZ, RFB) and (specification, SAO triples)
Procedure:
     Gather all knowledge element from all data pairs
     Create ID and node for each knowledge element, defining label, shape, colour, level and group
     Start loop:
         FOR each data pair:
         Fetch their corresponding ID,
         Add edge between the pair
     End loop
Output:
     Knowledge network

Figure 7. Procedure of establishing mapping and knowledge network construction.

and facilitate the comparison, expert analysis followed a similar methodology explained in Section 3.2. The time taken for analysing each patent is 20 min for Motivation, 30 min for Specification and 10 min for Structure. Expert analysis of mapping was not carried out and hence was not evaluated in this study. There are two main reasons for this. First, the task requires significantly more time to complete compared to other tasks. Second, unlike

Table 3. 15 patents from five engineering design invention categories selected for evaluation.

Invention categories	Patent Number	Notation
Bladeless fan (BF)	US-9599368-B2	BF1
	US-11396881-B2	BF2
	US-8454322-B2	BF3
Internal combustion engine (ICE)	US-8671920-B2	ICE1
	US-10247156-B2	ICE2
	US-7409933-B2	ICE3
Hybrid locomotive (HL)	US-8640629-B2	HL1
	US-7190133-B2	HL2
	US-9186997-B2	HL3
Hair dryer (HD)	US-10064469-B2	HD1
	US-8782920-B2	HD2
	US-10021952-B2	HD3
Folding bicycle (FB)	US-6986522-B2	FB1
	US-11535330-B2	FB2
	US-8473130-B2	FB3

fairly straightforward conceptualisation, each invention is understood by designers differently, therefore, challenging to accomplish objective mapping. As a result, the mapping outcome obtained in this study can be seen as a subjective interpretation of inventions by the method for designers to consider.

Figures 8–10 show a summary of how accurate the proposed method was in extracting motivation, specification and structure information for the 15 patents analysed. From the results, it can be seen that the method performed well in extracting motivation and specification. The majority of the results are close to expert analysis with a maximum difference of 5 in motivation for patent HL2. In almost all cases the number of results obtained using the proposed method are either equal to or more than expert analysis. This is particularly noticeable in Figure 10 in which 14 patents yield an additional number of results than expert analysis with the highest difference of 15. It can also be seen from Figure 10 that, unlike Figures 8 and 9, almost all structures extracted ended up with fewer numbers than the expert
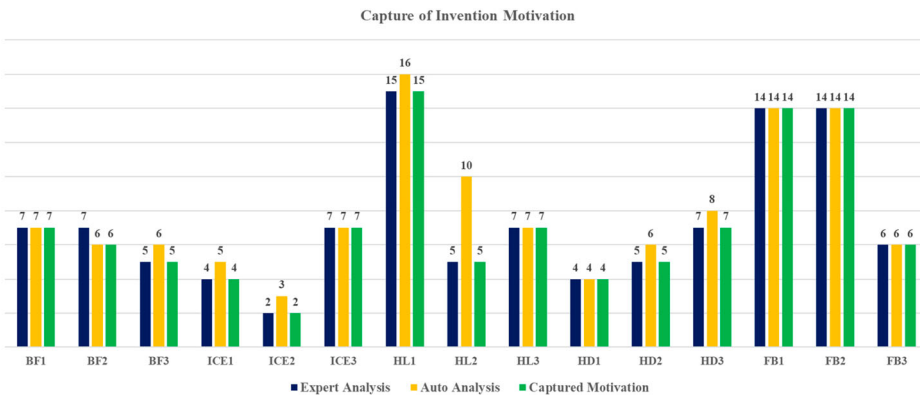

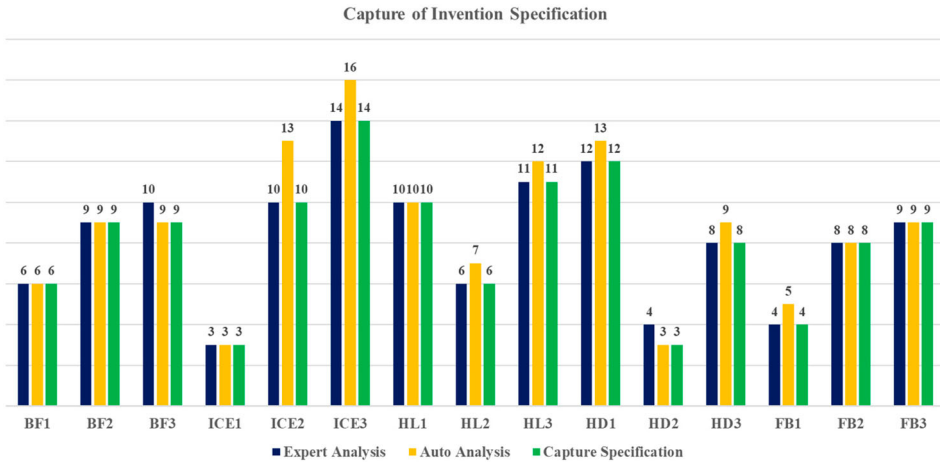**Figure 8.** Invention motivation extraction compared to expert analysis.

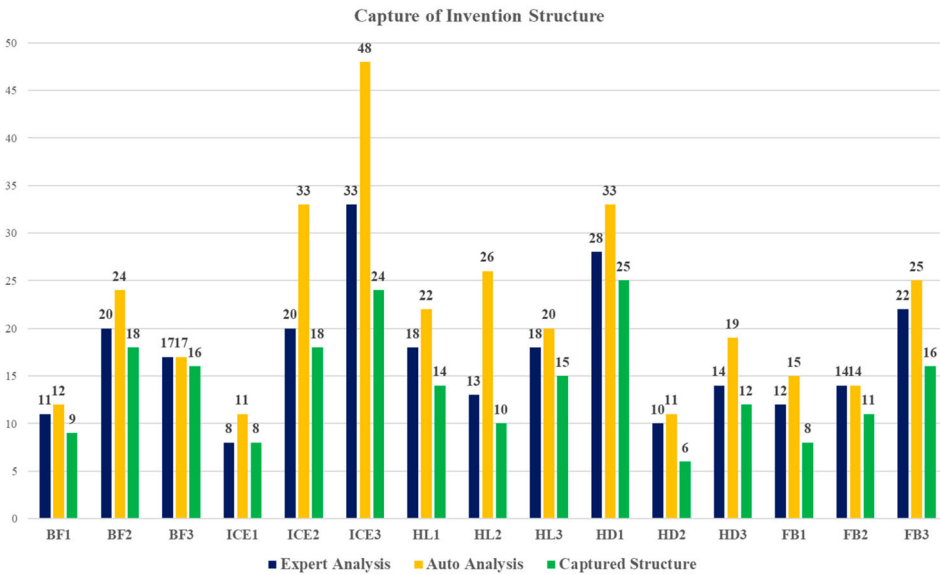**Figure 9.** Invention specification extraction compared to expert analysis.

Figure 10. Invention Structure extraction compared to expert analysis.

analysis. This is most likely due to the complex sentence structure used in claim sentences, suggesting a limitation of the proposed method.

As explained in Section 3.3, the extracted information is then conceptualised, more specifically, Motivation to TRIZ 39 parameters, Specification to RFB. Results obtained by the proposed method are evaluated against expert analysis and also Zero-Shot classification. Zero-Shot classification was carried out by using Transformers with a pre-trained model 'bart-large-mnli' from Hugging Face. The user-defined labels in each classification task correspond to TRIZ parameters, RFB function and RFB flow terms respectively. The results are shown in Figure 11–13. Results obtained by expert analysis are used as a reference for comparison.
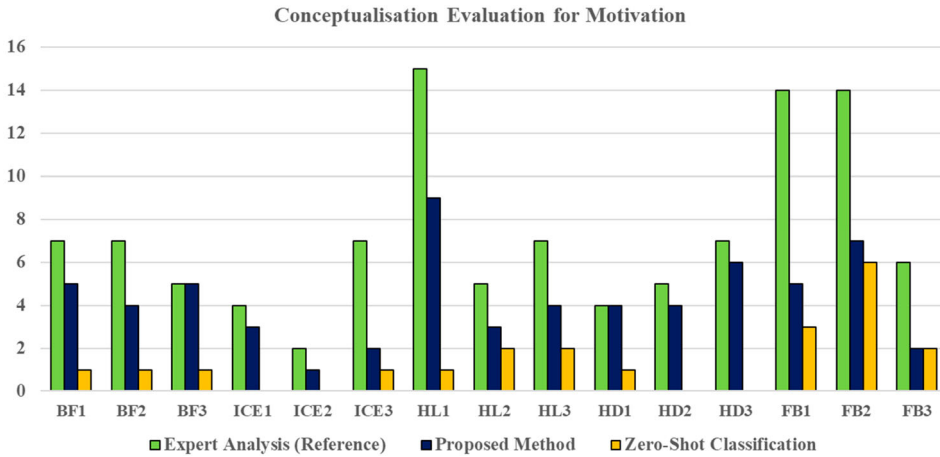
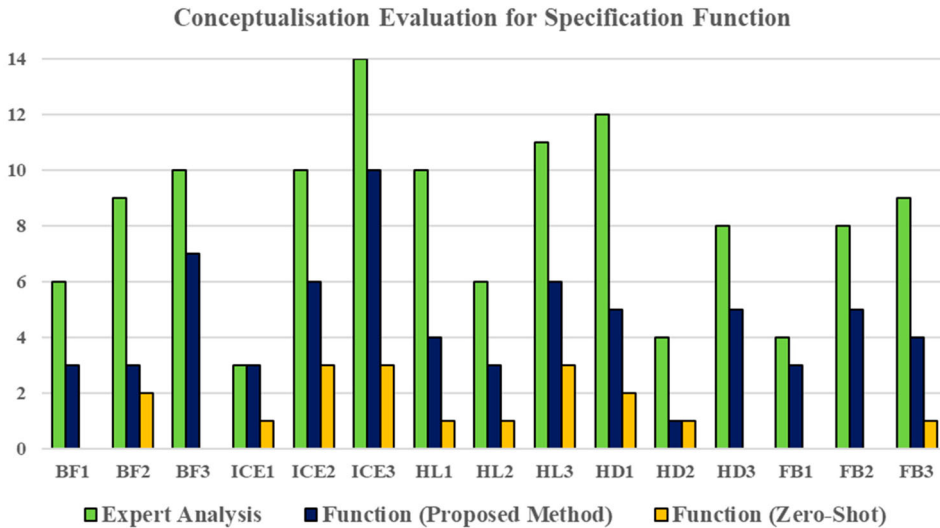**Figure 11.** Comparison of motivation conceptualisation.

Figure 12. Comparison of specification function conceptualisation.

The results indicate that the proposed method outperforms the Zero-Shot approach in almost every category. However, when compared to expert analysis it can be seen that there is still a noticeable difference even when the proposed method is applied.

With the aim of quantifying the performance of the proposed method, two performance measures are defined, namely Relative Performance (RP) and Absolute Performance (AP). Relative Performance only considers when there is a match between auto analysis and expert analysis, whilst Absolute Performance takes the 'noisy' data produced by the proposed method into consideration by having another ratio between the number of matched results and the total number of produced results. The formulas for these two measures can be found in Equations (1) and (2).
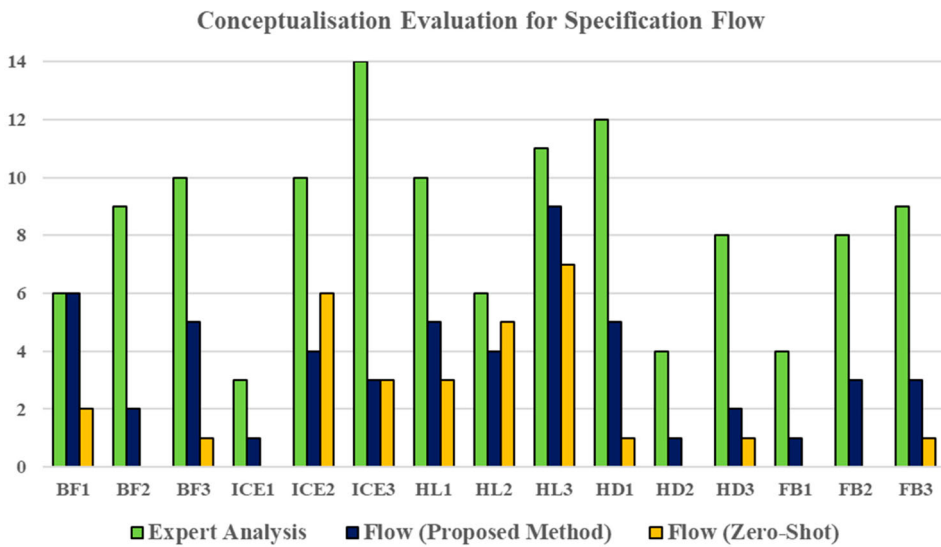
Figure 13. Comparison of specification flow conceptualisation.

$$\text{Relative Performance (RP) (\%)} = \frac{\text{No. of Matched Result}}{\text{No. of Expert Analysis}} \tag{1}$$

$$\text{Absolute Performance (AP) (\%)} = \text{RP (\%)} \times \frac{\text{No. of Matched Result}}{\text{No. of Auto Analysis}} \tag{2}$$

Table 4 shows a summary of performance measures for the method proposed in extracting and conceptualising invention knowledge. The values shown are the mean values obtained by taking the average of 15 patents. From the summary, it can be seen that the method performed very well in extracting invention motivation and specification, with a minimum of 89% match. Although 81% of the structure was captured successfully, a considerable amount of 'noisy' data was generated at the same time, resulting in a significant drop in absolute performance. As presented in Figure 11–13, the performance of the proposed method is far better than Zero-Shot classification. The accuracy of function conceptualisation is noticeably better than the flow, but all worse than motivation. The overall performance of the conceptualisation task suggests that further work needs to be done to improve the accuracy of summarising patent-specific knowledge into standardised engineering design vocabularies.

Table 4. Summary of method evaluation.

		Extraction		Conceptualisation			
		RP	AP	RP (Proposed Method)	AP (Proposed Method)	RP (Zero-Shot)	AP (Zero-Shot)
Motivation		99%	89%	64%	56%	18%	16%
Specification	Function	98%	91%	56%	52%	14%	13%
	Flow			44%	41%	22%	20%
Structure		81%	54%				

5. Prototype knowledge network

Based on the results obtained by analysing the 15 example patents, a prototype knowledge network is constructed using Pyvis following the procedure described in Section 3.3.2. Built around VisJS library, Pyvis features interactive environments that allow users to search, filter and drag knowledge elements within the network, with the aid of different levels and groups defined and keyword match, for example, see Figure 14. Other network construction toolkits can also be applied, in this study Pyvis was used as a demonstration.

Figure 15 shows an example use of the knowledge network. 'Loss of energy' is used as the search query, which returns the match invention motivation IDs. Detail description of the motivation can be obtained by cursor hovering. Figure 16 shows when 'Supply electrical
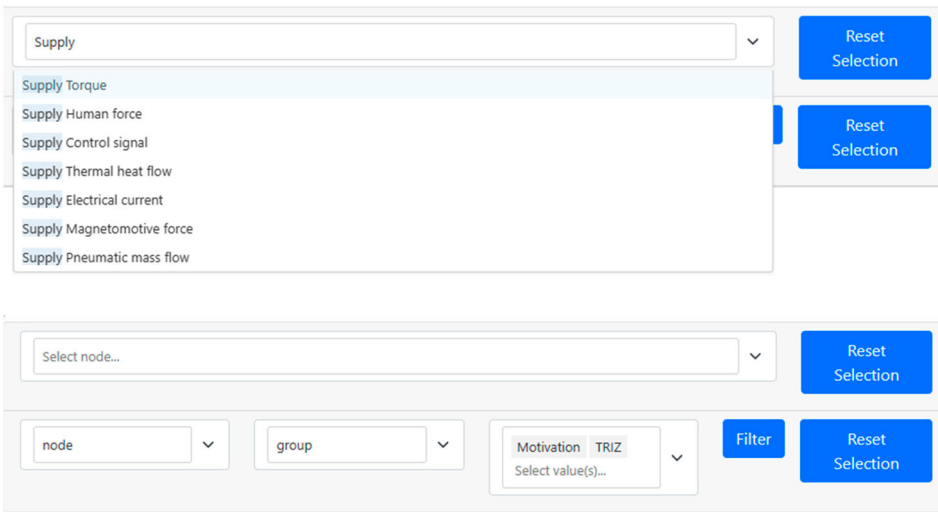

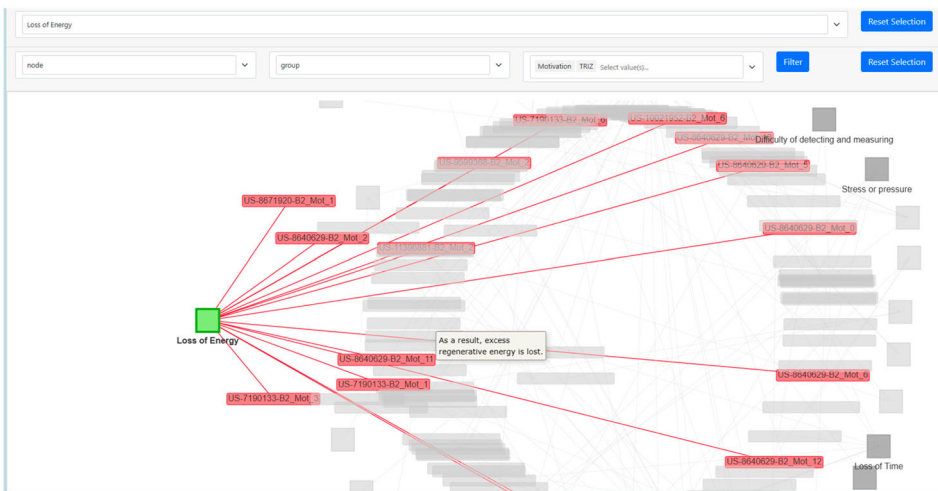**Figure 14.** Keyword search and filtering function offered by Pyvis.

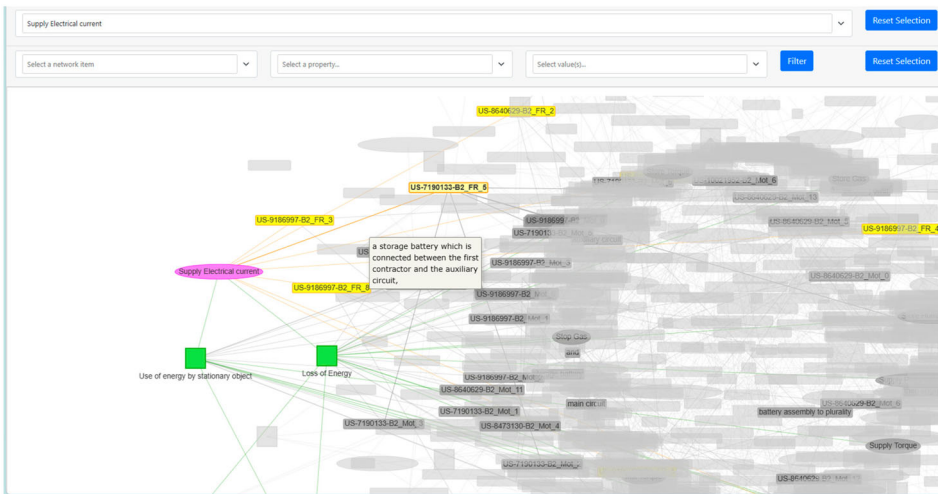**Figure 15.** Search results using 'Loss of Energy'.

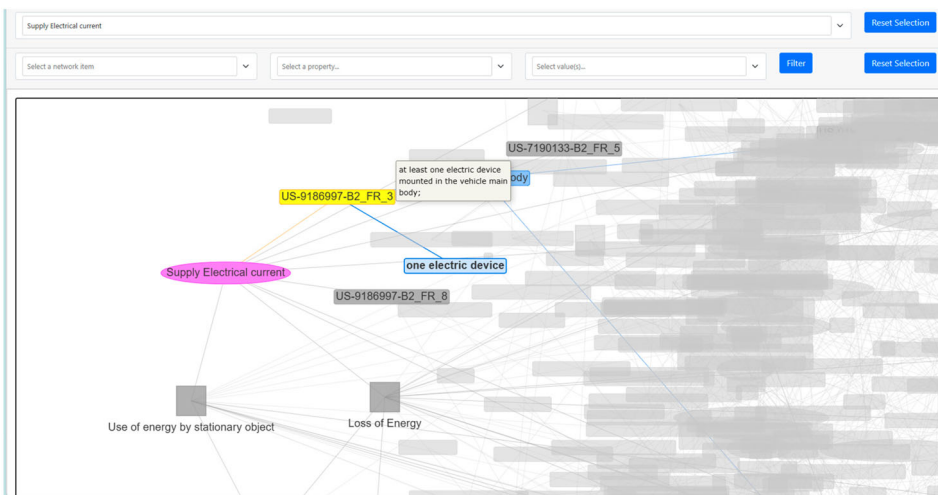**Figure 16.** Search results using 'Supply Electrical current'.

Figure 17. Results when a specific patent specification is selected.

current' is used as a search query. The related conceptualised motivations were also highlighted, 'Use of energy by stationary object' and 'Loss of energy', as well as related patent specification. The user can keep exploring the network by selecting the interested patent specification to reveal its relevant invention structure. For example, see Figure 17.

Figure 18 presents an overview of how the proposed knowledge network can be used in the early design stages. Designers can explore the knowledge base by sending queries concerning early design problems such as needs (Motivation), tasks (Specification) and designs (Structure). Connected knowledge for these queries will be highlighted to provide design insight. Designers are also able to trace down to specific patents for further investigation from the knowledge base.
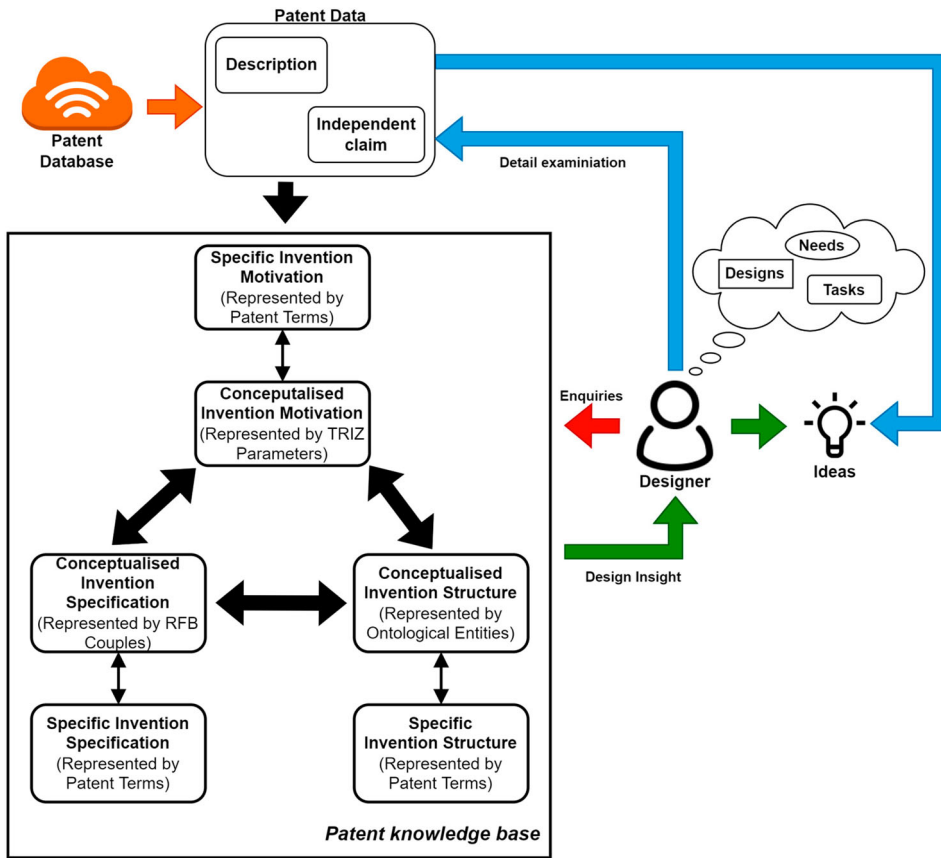
Figure 18. Overview of a linked patent knowledge base for early design use.

6. Discussion

The approach developed in this study provides a structured representation of invention knowledge that suits the natural way of designers' thinking; aiming to offer insight at early stages of design and support designers in making informed decisions prior to detail examination of patents. Drawn from the descriptive section of patents, the approach enables designers to quickly understand the need for an invention (Motivation), functions to be accomplished (Specification) and the conceptual solution (Structure), without the need for reading the patent document. In addition to the 'what', the linking between this knowledge is established in this study, providing insight into the 'how', e.g. how the need for an invention is associated with functions and then with conceptual solutions. This linked knowledge is then visualised using a network map as a convenient infographic.

With the use of TRIZ parameters and RFB terms as conceptualised motivation and specification, cross-patent analysis is enabled, leading to the establishment of a knowledge base. By looking at a number of analysed patents at the same time, designers can identify popular motivations from the inventions analysed and continue exploring the network to obtain detailed information regarding existing solutions. The example network established in this study can be navigated by sending queries to quickly locate target information, providing

intuitive usage. Compared to existing knowledge graphs and semantic networks developed in the literature, the knowledge base established in this study can provide traces back to individual patents. This can be beneficial to designers when detailed examination of patents is necessary. In other words, the knowledge base developed can be used as decision-making support tool to help designers to decide whether to investigate specific patents in detail in the early design stages.

In this study pre-trained models were used where possible to provide quick analysis. According to the evaluation results from the 15 sample patents analysed, the proposed approach performed well in extracting the invention knowledge. However, performance starts to drop when performance conceptualisation, suggesting the limitation of pre-trained models without fine-tuning. The authors believe that having user-labelled data will improve the performance to a great extent but it is impractical when cross-invention analysis is preferred as this will require a considerable amount of time and effort. Of course, if designers are willing to input more in the development and improvement of user-labelled datasets, they are likely to get a more accurate and sensible outcome. If they are looking for some quick yet also informative results, then standard ready-to-use techniques might still be sufficient. The main purpose of the approach developed in this study is to show the practicality of automatically identifying and linking invention Motivation, Specification and Structure from patent descriptive sections, aiming to provide quick insight into invention design knowledge for informed decisions. Therefore, the accuracy of the automatic analysis is not the primary concern here.

7. Conclusion

Being arguably the largest database of freely accessible engineering design information, patents can be seen as a great source of knowledge. However, it is commonly agreed that patents are difficult to understand even for professionals. As a result, it is beneficial to reformulate invention knowledge into a structured representation that suits the natural way of designer's thinking. In this paper, an approach is proposed, which aims to demonstrate how to identify and link invention Motivation, Specification and Structure in a short time using techniques such as natural language processing, sentiment analysis, sentence embeddings and semantic correlation. The approach focuses on the patent description and independent claim section, incorporating TRIZ 39 parameters and reconciled functional basis (RFB) to provide designers with a more comprehensive and systematic understanding with respect to engineering design. This approach enables a knowledge base to be established, existing as a network map in this study, offering great potential for designers to explore ideas at the early design stage. The designers are able to obtain insight into areas such as common motivations and specifications behind inventions and therefore make informed decisions. The knowledge base also offers traceability to individual patents, allowing designers to identify specific patents for further investigation.

Disclosure statement

ORCID

Pingfei Jiang http://orcid.org/0000-0001-6163-2632
Salvatore Sorce http://orcid.org/0000-0003-1976-031X

References

Abbas, A., L. Zhang, and S. U. Khan. 2014. "A Literature Review on the State-of-the-art in Patent Analysis." *World Patent Information* 37: 3–13. https://doi.org/10.1016/j.wpi.2013.12.006

Akerdad, M., A. Aboutajeddine, and M. Elmajdoubi. 2022. "Development of a Redesign Process Based on Reverse Engineering and Patent Circumvention." *International Journal on Interactive Design and Manufacturing (IJIDeM)* 17: 261–277.

Atherton, M., P. Jiang, D. Harrison, and A. Malizia. 2017. "Design for Invention: Annotation of Functional Geometry Interaction for Representing Novel Working Principles." *Research in Engineering Design* 29: 245–262. https://doi.org/10.1007/s00163-017-0267-2

Aurisicchio, M., R. H. Bracewell, and G. Armstrong. 2013. "The Function Analysis Diagram Intended Benefits and Coexistence with Other Functional Models." *AI EDAM* 27: 249–257. https://doi.org/10.1017/S0890060413000255.

Cao, G., P. Luo, L. Wang, and X. Yang. 2016. "Key Technologies for Sustainable Design Based on Patent Knowledge Mining." *Procedia CIRP* 39: 97–102. https://doi.org/10.1016/j.procir.2016.01.172

Cascini, G., and M. Zini. 2008. "IFIP International Federation for Information Processing." In *Computer Aided Innovation (CAI)*. Vol. 277, edited by G. Cascini, 31–42. Boston: Springer.

Chan, E. M., A.-L. Kor, K. W. Ng, M. C. Ang, and A. N. A. Wahab. 2021. "A Conceptual Design Framework Based on TRIZ Scientific Effects and Patent Mining." *International Journal of Advanced Computer Science and Applications* 12: 43–50. http://doi.org/10.14569/IJACSA.2021.0121206.

Chen, R. 2009. "Design Patent map Visualization Display." *Expert Systems with Applications* 36 (10): 12362–12374. https://doi.org/10.1016/j.eswa.2009.04.049

Cheng, S., J. Mi, R. Zhao, and X. Yang. 2016. "Patent Product Innovation Re-Design Based on Function Analysis." *Recent Patents on Mechanical Engineering* 9 (1): 71–78. https://doi.org/10.2174/22127976 09666151229194307.

Chiarello, F., P. Belingheri, and G. Fantoni. 2021. "Data Science for Engineering Design: State of the art and Future Directions." *Computers in Industry* 129: 1–17. https://doi.org/10.1016/j.compind.2021. 103447.

Cho, Y., Y. J. Han, J. Hwang, J. Yu, S. Kim, C. Lee, S. Lee, and K. P. Yi. 2021. "Identifying Technology Opportunities for Electric Motors of Railway Vehicles with Patent Analysis." *Sustainability* 13: 1–13. https://doi.org/10.3390/su13052424.

Dieter, G. E., and L. C. Schmidt. 2012. *Engineering Design*. 4th ed. McGraw Hill.

Esposito, M., E. Damiano, A. Minutolo, G. De Pietro, and H. Fujita. 2020. "Hybrid Query Expansion Using Lexical Resources and Word Embeddings for Sentence Retrieval in Question Answering." *Information Sciences* 514: 88–105. https://doi.org/10.1016/j.ins.2019.12.002

Fantoni, G., R. Apreda, F. Dell'orletta, and M. Monge. 2013. "Automatic Extraction of Function–Behaviour–State Information from Patents." *Advanced Engineering Informatics* 27 (3): 317–334. https://doi.org/10.1016/j.aei.2013.04.004

Fu, K., J. Murphy, M. Yang, K. Otto, D. Jensen, and K. Wood. 2015. "Design-by-analogy: Experimental Evaluation of a Functional Analogy Search Methodology for Concept Generation Improvement." *Research in Engineering Design* 26 (1): 77–95. https://doi.org/10.1007/s00163-014-0186-4

Gadd, K. 2011. *TRIZ for Engineers: Enabling Inventive Problem Solving*. Chichester: Wiley.

Grawe, M. F., C. A. Martins, and A. G. Bonfante. 2017. "Automated Patent Classification Using Word Embedding." 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA).

Habibi, M., L. Weber, M. Neves, D. L. Wiegandt, and U. Leser. 2017. "Deep Learning with Word Embeddings Improves Biomedical Named Entity Recognition." *Bioinformatics (oxford, England)* 33 (14): i37–i48. https://doi.org/10.1093/bioinformatics/btx228

Han, Y., and M. Moghaddam. 2021. "Analysis of Sentiment Expressions for User-Centered Design." *Expert Systems with Applications* 171: 1–11. https://doi.org/10.1016/j.eswa.2021.114604.

Han, X., D. Zhu, M. Lei, and T. Daim. 2021. "R&D Trend Analysis Based on Patent Mining: An Integrated use of Patent Applications and Invalidation Data." *Technological Forecasting and Social Change* 167: 1–11. https://doi.org/10.1016/j.techfore.2021.120691.

Henriques, R., A. Ferreira, and M. Castelli. 2022. "A Use Case of Patent Classification Using Deep Learning with Transfer Learning." *Journal of Data and Information Science* 7 (3): 49–70. https://doi.org/10.2478/jdis-2022-0015

Higham, K., M. Contisciani, and C. De Bacco. 2022. "Multilayer Patent Citation Networks: A Comprehensive Analytical Framework for Studying Explicit Technological Relationships." *Technological Forecasting and Social Change* 179: 1–17. https://doi.org/10.1016/j.techfore.2022.121628.

Hirtz, J., R. Stone, D. A. Mcadams, S. Szykman, and K. L. Wood. 2002. *A Functional Basis for Engineering Design Reconciling and Evolving Previous Efforts*. Washington: National Institute of Standards and Technology.

Howard, T. J., S. J. Culley, and E. Dekoninck. 2008. "Describing the Creative Design Process by the Integration of Engineering Design and Cognitive Psychology Literature." *Design Studies* 29 (2): 160–180. https://doi.org/10.1016/j.destud.2008.01.001

Hu, J., S. Li, Y. Yao, L. Yu, G. Yang, and J. Hu. 2018. "Patent Keyword Extraction Algorithm Based on Distributed Representation for Patent Classification." *Entropy (Basel)* 20: 1–19. https://doi.org/10.3390/e20020104.

Ipo. 2020. *Patent Factsheets: Claims*. Edited by Intellectual Property Office.

Jack, H. 2013. *Engineering Design, Planning, and Management*. Academic press.

Jeong, B., J. Yoon, and J.-M. Lee. 2019. "Social Media Mining for Product Planning: A Product Opportunity Mining Approach Based on Topic Modeling and Sentiment Analysis." *International Journal of Information Management* 48: 280–290. https://doi.org/10.1016/j.ijinfomgt.2017.09.009

Jiang, P., M. Atherton, D. Harrison, and A. Malizia. 2017. "Framework of mechanical design knowledge representations for avoiding patent infringement." 21st International Conference on Engineering Design, ICED 17, August 21–25, The University of British Columbia, Vancouver, Canada.

Jiang, P., M. Atherton, and S. Sorce. 2021. "Automated Functional Analysis of Patents for Producing Design Insight." Proceedings of the Design Society: International Conference on Engineering Design, August 16–20, Gothenburg, Sweden. 541–550.

Jiang, P., M. Atherton, S. Sorce, D. Harrison, and A. Malizia. 2018. "Design for Invention: A Framework for Identifying Emerging Design–Prior art Conflict." *Journal of Engineering Design* 29 (10): 596–615. https://doi.org/10.1080/09544828.2018.1520204

Jiang, S., J. Hu, C. L. Magee, and J. Luo. 2022a. "Deep Learning for Technical Document Classification." *IEEE Transactions on Engineering Management*, 1–17. https://doi.org/10.1109/TEM.2022.3152216.

Jiang, S., and J. Luo. 2022. "Technology Fitness Landscape for Design Innovation: A Deep Neural Embedding Approach Based on Patent Data." *Journal of Engineering Design* 33 (10): 716–727. https://doi.org/10.1080/09544828.2022.2143155.

Jiang, S., S. Sarica, B. Song, J. Hu, and J. Luo. 2022b. "Patent Data for Engineering Design: A Critical Review and Future Directions." *Journal of Computing and Information Science in Engineering* 22: 1–13. https://doi.org/10.1115/1.4054802.

Kang, J., J. Kim, J. Lee, S. Park, and D. Jang. 2015. "A Methodology for Patent Function Analysis." In International Conference on Economics and Business Management (EBM-2015), July 29–30, Phuket, Thailand.

Kim, Y. G., J. H. Suh, and S. C. Park. 2008. "Visualization of Patent Analysis for Emerging Technology." *Expert Systems with Applications* 34 (3): 1804–1812. https://doi.org/10.1016/j.eswa.2007.01.033

Koh, E. C. Y. 2020. "Read the Full Patent or Just the Claims? Mitigating Design Fixation and Design Distraction When Reviewing Patent Documents." *Design Studies* 68: 34–57. https://doi.org/10.1016/j.destud.2020.02.001

Krestel, R., R. Chikkamath, C. Hewel, and J. Risch. 2021. "A Survey on Deep Learning for Patent Analysis." *World Patent Information* 65: 1–13. https://doi.org/10.1016/j.wpi.2021.102035.

Li, Z., M. Atherton, and D. Harrison. 2014. "Identifying Patent Conflicts: TRIZ-Led Patent Mapping." *World Patent Information* 39: 11–23. https://doi.org/10.1016/j.wpi.2014.07.002

Li, Z., Y. Fan, B. Jiang, T. Lei, and W. Liu. 2019b. "A Survey on Sentiment Analysis and Opinion Mining for Social Multimedia." *Multimedia Tools and Applications* 78 (6): 6939–6967. https://doi.org/10.1007/s11042-018-6445-z.

Li, M., X. Ming, L. He, M. Zheng, and Z. Xu. 2015. "A TRIZ-Based Trimming Method for Patent Design Around." *Computer-Aided Design* 62: 20–30. https://doi.org/10.1016/j.cad.2014.10.005

Li, Z., and D. Tate. 2019. "Automatic Function Interpretation: Using Natural Language Processing on Patents to Understand Design Purposes." Proceedings of the ASME 2010 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference, August 15–18, Montreal, Quebec, Canada.

Li, Z., D. Tate, C. Lane, and C. Adams. 2012. "A Framework for Automatic TRIZ Level of Invention Estimation of Patents Using Natural Language Processing, Knowledge-Transfer and Patent Citation Metrics." *Computer-Aided Design* 44 (10): 987–1010. https://doi.org/10.1016/j.cad.2011.12.006

Li, H., J. Yuan, R. Tan, and Q. Peng. 2019a. "Design Around Bundle Patent Portfolio Based on Technological Evolution." *Chinese Journal of Mechanical Engineering* 32 (1): 1–16. https://doi.org/10.1186/s10033-019-0400-4.

Liu, H., W. Li, and Y. Li. 2021. "A new Computational Method for Acquiring Effect Knowledge to Support Product Innovation." *Knowledge-Based Systems* 231: 1–24. https://doi.org/10.1016/j.knosys.2021.107410.

Liu, L., Y. Li, Y. Xiong, and D. Cavallucci. 2020. "A new Function-Based Patent Knowledge Retrieval Tool for Conceptual Design of Innovative Products." *Computers in Industry* 115: 1–16. https://doi.org/10.1016/j.compind.2019.103154.

Luo, J., S. Sarica, and K. L. Wood. 2019. "Computer-aided design ideation using InnoGPS." Proceedings of the ASME 2019: International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, August 18–21, Anaheim, CA, USA.

Luo, J., S. Sarica, and K. L. Wood. 2021. "Guiding Data-Driven Design Ideation by Knowledge Distance." *Knowledge-Based Systems* 218: 1–15. https://doi.org/10.1016/j.knosys.2021.106873.

Luo, J., B. Song, L. Blessing, and K. Wood. 2018. "Design Opportunity Conception Using the Total Technology Space map." *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 32 (4): 449–461. https://doi.org/10.1017/S0890060418000094

Mann, D. 2001. "An Introduction to TRIZ The Theory of Inventive Problem Solving." *Creativity and Innovation Management - TRIZ* 10 (2): 123–125.

Marley, M. 2015. "Full-text Patent Searching on Free Websites." *Business Information Review* 31: 226–236. https://doi.org/10.1177/0266382114564265

Miller, G. A. 1995. "WordNet." *Communications of the ACM* 38: 39–41.

Murphy, J., K. Fu, K. Otto, M. Yang, D. Jensen, and K. Wood. 2014. "Function Based Design-by-Analogy: A Functional Vector Approach to Analogical Search." *Journal of Mechanical Design* 136 (10): 1–16. https://doi.org/10.1115/1.4028093.

Qu, Y. B., P. H. Liu, W. Song, L. Z. Liu, and M. M. Cheng. 2020. "A Text Generation and Prediction System: Pre-training on New Corpora Using BERT and GPT-2." Proceedings of 2020 IEEE 10th International Conference on Electronics Information and Emergency Communication (ICEIEC 2020), 323–326.

Reimers, N., and I. Gurevych. 2019. "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks." Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, November 3–7, Hong Kong, China.

Ríos-Zapata, D., R. Duarte, J. Pailhès, R. Mejía-Gutiérrez, and M. Mesnard. 2017. "Patent-based Creativity Method for Early Design Stages: Case Study in Locking Systems for Medical Applications." *International Journal on Interactive Design and Manufacturing (IJIDeM)* 11 (3): 689–701. https://doi.org/10.1007/s12008-016-0352-1.

Roca-Gonzalez, J. L., J. A. Vera-Lopez, and G. Rodriguez-Bermudez. 2018. "Analysis of Patent #US2014/0319274A1: A Case Study of Simulations for New Designs Review." *International Journal of Simulation Modelling* 17 (3): 405–418. https://doi.org/10.2507/IJSIMM17(3)433

Sarica, S., J. Han, and J. Luo. 2023. "Design Representation as Semantic Networks." *Computers in Industry* 144: 1–11. https://doi.org/10.1016/j.compind.2022.103791.

Sarica, S., J. Luo, and K. L. Wood. 2020. "TechNet: Technology Semantic Network Based on Patent Data." *Expert Systems with Applications* 142: 1–15. https://doi.org/10.1016/j.eswa.2019.112995.

Sarica, S., B. Song, J. Luo, and K. L. Wood. 2021. "Idea Generation with Technology Semantic Network." *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 35 (3): 265–283. https://doi.org/10.1017/S0890060421000020

Siddharth, L., L. Blessing, and J. Luo. 2022a. "Natural Language Processing in-and-for Design Research." *Design Science* 8: 1–76. https://doi.org/10.1017/dsj.2022.16.

Siddharth, L., L. T. M. Blessing, K. L. Wood, and J. Luo. 2022b. "Engineering Knowledge Graph from Patent Database." *J. Comput. Inf. Sci. Eng* 22: 1–37. https://doi.org/10.1115/1.4052293.

Siddharth, L., G. Li, and J. Luo. 2022c. "Enhancing Patent Retrieval Using Text and Knowledge Graph Embeddings: A Technical Note." *Journal of Engineering Design* 33 (8-9): 1–22. https://doi.org/10.1080/09544828.2022.2144714.

Song, B., J. Luo, and K. Wood. 2019. "Data-Driven Platform Design: Patent Data and Function Network Analysis." *Journal of Mechanical Design* 141 (2): 1–37. https://doi.org/10.1115/1.4042083.

Soong, H.-C., N. B. A. Jalil, R. K. Ayyasamy, and R. Akbar. 2019. "The Essential of Sentiment Analysis and Opinion Mining in Social Media: Introduction and Survey of the Recent Approaches and Techniques." 2019 IEEE 9th Symposium on Computer Applications & Industrial Electronics (ISCAIE), April 27–28, Malaysia.

Sorce, S., A. Malizia, V. Gentile, P. Jiang, M. Atherton, and D. Harrison. 2019. "Evaluation of a Visual Tool for Early Patent Infringement Detection During Design." In *End-User Development*, edited by A. Malizia, S. Valtolina, A. Morch, A. Serrano, and A. Stratton. Cham: Springer. https://doi.org/10.1007/978-3-030-24781-2_12.

Sorce, S., A. Malizia, P. Jiang, M. Atherton, and D. Harrison. 2018. "A Novel Visual Interface to Foster Innovation in Mechanical Engineering and Protect from Patent Infringement." IOP Conf. Series: Journal of Physics: Conf. Series 1004.

Speer, R., J. Chin, and C. Havasi. 2017. ConceptNet 5.5: An Open Multilingual Graph of General Knowledge. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17).

Stein, R. A., P. A. Jaques, and J. F. Valiati. 2019. "An Analysis of Hierarchical Text Classification Using Word Embeddings." *Information Sciences* 471: 216–232. https://doi.org/10.1016/j.ins.2018.09.001

Thorne, C., and S. Akhondi. 2020. Word Embeddings for Chemical Patent Natural Language Processing. arXiv:2010.12912v1.

Valverde, U. Y., J.-P. Nadeau, and D. Scaravetti. 2017. "A new Method for Extracting Knowledge from Patents to Inspire Designers During the Problem-Solving Phase." *Journal of Engineering Design* 28 (6): 369–407. https://doi.org/10.1080/09544828.2017.1316361

Vasantha, G., J. Corney, R. Maclachlan, and A. Wodehouse. 2017. "The Analysis and Presentation of Patents to Support Engineering Design." Design Computing and Cognition'16. 209-226.

Wang, Y., S. Liu, N. Afzal, M. Rastegar-Mojarad, L. Wang, F. Shen, P. Kingsbury, and H. Liu. 2018. "A Comparison of Word Embeddings for the Biomedical Natural Language Processing." *Journal of Biomedical Informatics* 87: 12–20. https://doi.org/10.1016/j.jbi.2018.09.008

Wodehouse, A., G. Vasantha, J. Corney, A. Jagadeesan, and R. Maclachlan. 2018. "Realising the Affective Potential of Patents: A new Model of Database Interpretation for User-Centred Design." *Journal of Engineering Design* 29 (8-9): 484–511. https://doi.org/10.1080/09544828.2018.1448056

Wodehouse, A., G. Vasantha, J. Corney, R. Maclachlan, and A. Jagadeesan. 2017. "The Generation of Problem-Focussed Patent Clusters: A Comparative Analysis of Crowd Intelligence with Algorithmic and Expert Approaches." *Design Science* 3: 1–31. https://doi.org/10.1017/dsj.2017.19.

Yan, B., and J. Luo. 2017. "Measuring Technological Distance for Patent Mapping." *Journal of the Association for Information Science and Technology* 68 (2): 423–437. https://doi.org/10.1002/asi.23664

Yang, W., G. Cao, Q. Peng, J. Zhang, and C. He. 2022. "Effective Identification of Technological Opportunities for Radical Inventions Using International Patent Classification: Application of Patent Data Mining." *Applied Sciences* 12: 1–24. https://doi.org/10.3390/app12136755.

Yang, S.-Y., and V.-W. Soo. 2012. "Extract Conceptual Graphs from Plain Texts in Patent Claims." *Engineering Applications of Artificial Intelligence* 25 (4): 874–887. https://doi.org/10.1016/j.engappai.2011.11.006

Yin, W., J. Hay, and D. Roth. 2019. Benchmarking Zero-shot Text Classification: Datasets, Evaluation and Entailment Approach. arXiv:1909.00161.

Yuan, L., Y. Liu, Z. Sun, Y. Cao, and A. Qamar. 2016. "A Hybrid Approach for the Automation of Functional Decomposition in Conceptual Design." *Journal of Engineering Design* 27 (4-6): 333–360. https://doi.org/10.1080/09544828.2016.1146237

Zuo, H., Q. Jing, T. Song, L. Sun, P. Childs, and L. Chen. 2022. "WikiLink: An Encyclopedia-Based Semantic Network for Design Creativity." *Journal of intelligence* 10 (4): 1–24. https://doi.org/10.3390/jintelligence10040103.

Zvarevashe, K., and O. O. Olugbara. 2018. "A Framework for Sentiment Analysis with Opinion Mining of Hotel Reviews." 2018 Conference on Information Communications Technology and Society (ICTAS).