Sustainable Product Innovation using Patent Mining and TRIZ

Chun Kit Chan[1], Kok Weng Ng[1], Mei Choo Ang[2], Chin Yuen Ng[1], and Ah-Lian Kor[3]

[1] School of Mechanical, Materials and Manufacturing Engineering, University of Nottingham Malaysia, Malaysia
[2]Insitute of IR 4.0, Universiti Kebangsaan Malaysia, Bangi, Malaysia
amc@ukm.edu.my
[3]School of Built Environment, Engineering, and Computing, Leeds Beckett University, Leeds, United Kingdom

Abstract. Sustainable issues have become more serious due to the rapid development of the global economy. Sustainable design is an approach for designing or creating sustainable products/solutions based on sustainable development principles. Patent documents contain a lot of useful inventive information which will be useful for sustainable product design. However, they are dense and lengthy due to excessive overladen technical terminology. Automatic text mining tool in patent analysis is therefore, in great demand to assist innovators or patent engineers in their patent search. The main focus of this work is to develop a patent mining prototype to extract sustainable design information from the patent database and recommend potential solutions to the user by using Patent Mining and TRIZ. The TRIZ problem solving process and details of patent mining will be described in this paper. Patent mining techniques include tokenization, stop words filtering, stemming, lemmatization and classification. The patent mining techniques were implemented together with relevant sustainable design indicators to identify patent documents that contained the most relevant sustainable design solution or suggestions. A sustainable design problem is illustrated in this paper to demonstrate how a TRIZ user can utilize the implemented patent mining techniques and sustainable design indicators to obtain a sustainable solution for a design problem.

Keywords: Sustainable, Product Innovation, Patent Mining, TRIZ

1 Introduction

With rapid development of the global economy, issues such as resource shortages, air pollution , etc., have become more prevalent [1, 2] and are major causes for sustainability-related problems. It has been suggested that AI be employed for sustainability and innovation [3] to support UN 2030 sustainable development goals [4]. Sustainable product innovation (to support sustainable development) is necessary to effectively address such sustainability problem. Sustainable development includes ecological design, green or environmental design, and sustainable design [5]. Sustainable design is the higher level of green and environmentally friendly design evolution with an end goal developing a sustainable product or solution [2]. The entire process encompasses a systematic adaptation or embedment of sustainable development principles [6]. However, challenges faced by designers stem from the lack of suitable design tools during conceptual design phase [2, 7].

Theory of Inventive Problem Solving (TRIZ) is an innovative problem-solving tool that fosters a systematic study of patterns of invention in the global patent literature [8]. It eliminates the need for compromise and trade-off caused by conflicts as well as contradictions amongst different performance measures. Genrich Altshuller developed TRIZ in Russia [9] by analyzing huge amounts of patent documents because he recognized the fact that ideas of invention and new concepts were in published inventions. Product innovation connotes a solved problem that we are trying to solve.

Patent analysis [10-14] in high-tech management has become more prominent as the innovation process and innovation cycle becomes more complex and shorter, and therefore, resulting in an unpredictable and unstable market demand. Patents contain a lot of technical information which is a source for technological-scientific innovations. However, they are dense and lengthy due to verbose technical terminology and details. Such documents requires extensive as well as intensive manual analysis. The method of reading or scanning the indexed patent documents to extract information from a long list of unprocessed results is a very time-consuming and is not a trivial process that involves careful manual selection. Data mining can be used to address this problem. Data mining is a system that automatically extract useful information from massive databases. Text mining of patent information or patent mining is similar to data mining, but for the full-text patent analysis, it specifically extracts useful knowledge from patent documents which typically comprises poorly structured texts [15]. Automatic text mining tool in patent analysis is therefore, in great demand to assist innovators or patent engineers.

A patent mining prototype is developed in this work to extract sustainability design information from a patent database and recommends Patent Mining and TRIZ solutions to potential users. . The focus would be primarily on the methodology of extracting textual information from patent documents using Python. In this paper, the TRIZ problem solving process, and patent mining which mainly focus on text mining process will be introduced in Section 1 and Section 2. The details of the implemented patent mining process such as tokenization, filtering, stemming and lemmatization and classification will be elaborated in Section 3. In section 4, sample results of the sustainable design recommendations will be displayed and discussed. This paper demonstrates how TRIZ can be applied together with patent mining in a consumer product example in Section 5. Finally, this paper is concluded with some future work in Section 6.

2 TRIZ Problem Solving Process

The general process of solving an engineering problem is shown in Fig. 1. Engineers utilize all their knowledge and expertise to translate an engineering problem (specific problem) to a specific solution. In applying TRIZ problem solving, the TRIZ engineering contradiction is applied by defining an engineering problem in the form of contradicting features (general problem) which will then propose a list of possible general solutions in the form of inventive principles which are derived from learning repeating patterns of problems and solutions from patent information. From the suggested inventive principles which are actually general solution, the engineer will have to translate these inventive principles into specific solutions for the engineering problem as shown in Fig. 2. However, it is a not an easy task to define the contradiction features i.e. map a specific problem to 39 improving and worsening features [7]. In other words, it remains a very challenging task for an engineer to translate the inventive principles (general solutions) to specific solutions. However, with some general ideas from the inventive principles, the task of finding a specific solution based on the inventive principles is relatively easier compared to the problem solving process without TRIZ. However, this approach might not work in cases when there are contradictions or conflicts that are difficult to be resolved before the generation of good solutions [16].

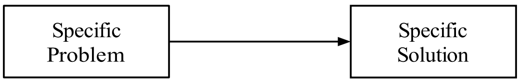


Fig. 1. The general process of problem solving without TRIZ.

For the TRIZ problem solving approach shown in Fig. 2, engineers need to analyze a specific problem encountered. Subsequently, they map the problem into a general problem in TRIZ. The general problem will have some general TRIZ solutions generated through the application of TRIZ tools including contradiction matrix, substance-field modelling, and even ARIZ (a Russian acronym for the TRIZ tool, "Algorithm for Inventive Problem Solving"). However, it is common for engineers to struggle to translate general TRIZ solutions to specific solutions as the general TRIZ solutions are in the form of inventive principles which are very abstract and require in-depth domain knowledge and expertise. Based on this identified gap, in this research, a patent mining system has been developed to assist engineers in the translation of general TRIZ solutions. Patents related to the inventive principles will be identified and further manually examined to provide more specific ideas to engineers.

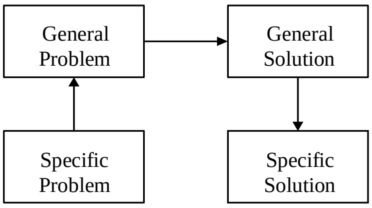


Fig. 2. The general process of problem solving with TRIZ.

Thus, with the help of patent mining as illustrated in Fig. 3, engineers can speed up the process of devising specific solutions by reviewing relevant patent documents in a shorter time span before deciding on which patent they should further examine in detail. By incorporating patent mining in the TRIZ engineering contradiction framework, it will facilitate an engineer's task of designing a specific solution for an engineering problem. Undeniably, the work entailed will be much easier compared to manual individual patent search or standalone TRIZ engineering contradiction tools. Although TRIZ tools are not specifically created for sustainable design, these tools can be applied to sustainable design if engineers applied the inventive principles grounded on sustainability.

Thus, with the incorporation of patent mining, our proposed patent mining system can assist engineers consider sustainability in their specific solutions based on the recommended inventive principles. This is because patent mining is a computational approach that can facilitate a faster and more focused search of large patent databases. Large amount of relevant and specific information (e.g. information related to sustainability or specific solutions) could be retrieved easily and quickly to support engineers in their sustainable solution designs.

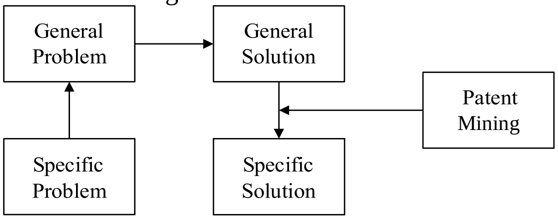


Fig. 3. TRIZ problem solving process with the help of Patent Mining.

3 Patent Mining

Patent documents are divided into two categories namely structured and unstructured data [12]. Structured data refers to information that are well organized such as patent number, assignees or filing date [11]. For unstructured data, they are not well organized and normally, they are full texts of various lengths and contains, such as Title, Abstract, Claims and Descriptions [11]. Abstract, Claims and Descriptions contain useful information such as the technical features of an invention.

There are 3 main stages of Patent Mining for this research work (see Fig. 4). The first stage is patent information extraction, which includes downloading related patent documents from USPTO and extract unstructured data from all the documents. After extracting all useful unstructured data from the patent document, information that are related to sustainability will be further extracted. Relevant sustainability information will be fed into the design conceptualization process. The detail patent mining process expanded from these main stages is shown in Fig. 5.

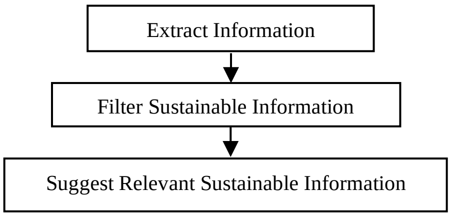

Extract Information
Filter Sustainable Information
Suggest Relevant Sustainable Information

Fig. 4. Patent Mining Framework.

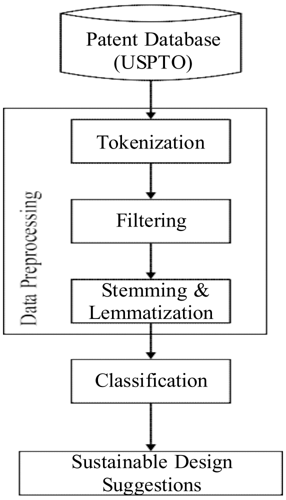


Fig. 5. Patent Mining Workflow.

3.1 Extraction of Patent Information

The data source is patent information which is a key element of patent data analysis. Care must be taken to ensure a sizeable dataset viewing the fact that large amount of data could be messy. Therefore, it is necessary to provide scope and boundaries of the research, only relevant design criteria and context of interest are considered. Next, patent documents are retrieved from suitable open online patent databases such as USPTO, Google Patent, SIPO, EPO, etc..

In this work, patent information that are related to sustainability and design is mainly obtained from USPTO. This is facilitated by developing an automated search program for USPTO advanced search facility. The information that is retrieved from USPTO is in a HTML format, where only title, abstract, claims and description are extracted and stored in a CSV format. Subsequently, the patent document is imported into a Python workspace and store in a dictionary format with data for the patent number, title, and abstract. A total of 22281 patent documents related to sustainability have been extracted in this research work. These patent documents are stored as csv files for further search based on sustainable design indicators. Processes encompass tokenization, stopwords filtering, and stemming as well as lemmatization.

3.2 Data Pre-processing

Extracted information from the unstructured data from all relevant patent documents are lengthy. Therefore, data pre-processing is required to reduce the size of the dataset to transform them to structured data for easy analysis. In this research work, there are 3 main processes in the data pre-processing, namely Tokenization, Filtering and Stemming as well as Lemmatization.

Natural Language Toolkit (NLTK) is a Python library that work with human language data for the application in statistical Natural Language Processing (NLP) [17]. It contains text processing libraries for parsing, tokenization, stemming, semantic reasoning, classification and tagging. Therefore, the imported patent information which is stored in the Python dictionary will undergo pre-processing before classification. This is to remove unrelated information from the corpus and improve the search results through the extraction of more relevant design suggestions.

A. Text Segmentation (Tokenization). Tokenization is one of the text mining techniques that is used to split text into smaller units which is known as tokens. The non-text characters in the patent information such as tabs, punctuation, etc. are removed in this process. Tokens can be individual word, phrases or even a whole sentence. The unstructured data are segmented into smaller units for summarization [11]. Additionally, the tokens can be further processed by filtering process.

NLTK provides a tokenization module which can be used to divide a text into tokens which are “word_tokenize()” and “sent_tokenize()”. Word_tokenize() is used to break the text data into each word with punctuation. Sent_tokenize() is used to partition the text data into sentences. In this work, sent_tokenize() has been used to split the patent information. This is to present the result of a potential solution in the form of a sentence instead of some single words because a sentence will make more sense than words.

B. Stopwords Filtering. Filtering is a process to remove words from a document. Term Frequency – Inverse Document Frequency (TF-IDF) is a numerical statistic that indicates the importance of a word which relate to a text document or corpus and is widely used in text filtering process. Words with low TF and IDF are removed in indexing of a collection. However, using TF-IDF alone does not prevent undesirable words such as function words from being calculated. Therefore, stop words filtering is added into the process.

Stopwords are words that do not contain any significant meaning to search queries and stopwords are commonly used in English, such as “as, the, be, are” etc. Stopwords are normally filtered out before the processing of natural language data in operation because they appear too frequently in the patent text and lose their purpose as search terms.

Each programming language will provide its own list of stopwords for deployment because there is no universal list of stop words used by all the natural language processing tools and not all the tools use that list at all. USPTO has its own list of stopwords which can be found in USPTO official website. NLTK has its own list of stopwords too which can be imported from “nltk.corpus”.

C. Stemming and Lemmatization. In a full text of extracted information, there are words, which has similar meaning but in different form, which need to be reduced to its base (root) form. For example, the word “play” is the root form of “plays”, “playing” and “played”. This could be problematic for text data analysis, and it can be solved by applying stemming and lemmatization.

Stemming is a process of removing affixes from the word and transforming the word into its root form. Sometimes the root is not an actual word but might be part of the word. Lemmatization is a process of reducing the inflected words to its actual base form of word which is known as the Lemma. Lemma is the root word in lemmatization. WordNet Lemmatizer is a Python NLTK package that is used to lookup lemmas of words using the WordNet database.

Lemmatization is used in this work because it returns the actual word, which is more accurate than stemming. However, lemmatization has consumed higher processing time compared to stemming. WordNet Lemmatizer can be imported from NLTK to the workspace using “nltk.stem”.

3.3 Classification

Study on sustainable design indicators is very important because it can render design suggestions more related to the sustainability problem. Sustainable design indicators have been classified into three groups which are environmental, economic, and social. Sustainable design indicators are used in the extraction process to extract the most similar sentence and classified into these 3 groups.

A. Sustainable Design Indicators. In order to obtain suggestions that are related to sustainable design, relevant design indicators have been extracted from literature and articles. Sustainable design indicators selected for this research work is shown in Table 1 [5].

Table 1. Sustainable Design Indicators [5].

Environmental	**Economic**	**Social**
High material utilization	Low Production Costs	Practicability
Energy and resource conservation	Low Transportation Costs	High Degree of Intelligence
Biodegradable Material	Low Recovery Costs	Security
Material Easy Recovery	Low Maintenance Costs	High Reliability
Identifiable Material	Low Using Costs	Ease of Use
Material of low pollution	High Economic Benefits	High Capacity Utilization

B. Gensim. Gensim (Generate Similarity) is a topic detection modelling which is a technique that is very apt for this work. It determines the similarity between a pair of documents or at least two sets documents. Cosine Similarity and Vector Space modelling are used in Gensim.

Vector Space modelling is a model for representing text document as vectors [18]. Bag-of-Words is a document representation which is used to convert document to vectors as depicted in Fig. 6. After pre-processing, the documents will retain the meaningful words with the corpus converted to vectors for future use. All the words in the Bag-of-Words are allocated a specific integer ID as illustrated in Fig. 7.

[‘a’, ‘b’, ‘c’, ‘d’, ‘e’, ...]

Fig. 6. Example of Bag-of-Words.

[‘a’: 0, ‘b’: 1, ‘c’: 2, ‘d’: 3, ‘e’: 4, ...]

Fig. 7. Unique ID assigned to all words in Bag-of- Words.

the function “doc2bow()” is used to convert the tokenized documents to vectors. This function calculates the number of occurrences of every word before transforming a word into its integer word id and a sparse vector is returned as a result.

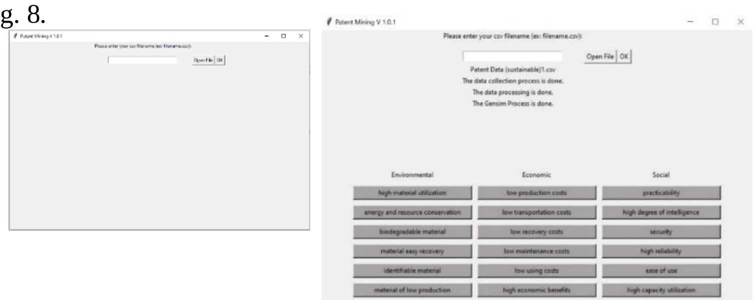
The sparse vector is used as a representation of a document and it assumes the form (“word id”, “Occurrences”). Therefore, it is read as the word “a” (id 0) and word “b” (id 1) appearing once.

Cosine similarity is used to measure and make comparisons of documents similarity or ascribe a rank to the documents with respect to a given vector of query words [19]. The Cosine similarity approach involves the computation of the cosine of the angle between vectors x and y. The similarity between two term-frequency vectors can be calculated using the equation 1 below.

$$sim(x, y) = \frac{x.y}{|x||y|} \tag{1}$$

where |*x*| = The Euclidean distance of vector x; |*y*| = The Euclidean distance of vector y.

In Cosine similarity, when both the vectors are orthogonal (i.e. at 90 degrees to each other) then the cosine value is zero, which implies they are not a match. If the angle is very small, the cosine value will be very close to 1, which means the vectors have greater match [17]. To prepare for similarity queries, all collected data documents are imported into Python workspace to compare against subsequent queries. A user types in the query and uses it to search for a related sentence by comparing it against a pool of collated data documents. To reiterate, an initial search based on the word ‘sustainability’ yields a search results of 22281 patent documents. Next, sustainable design indicators are used as query inputs for further search on this pool of retrieved documents. The final search results are grouped under the sustainable design indicators as shown in Fig. 8.



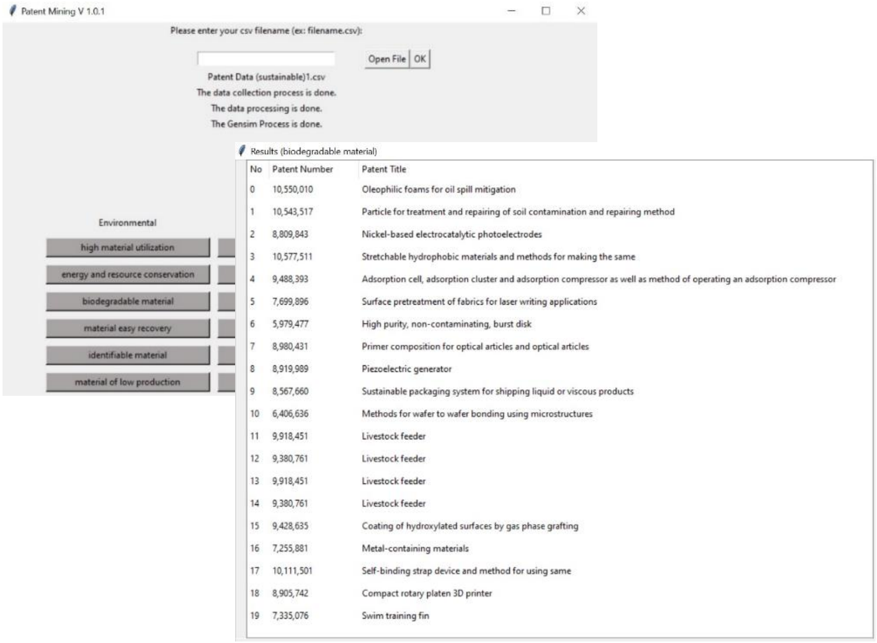
Please enter your csv filename (ex: filename.csv):

Please enter your csv filename (ex: filename.csv):

Patent Data (sustainable)1.csv
The data collection process is done.
The data processing is done.
The Gensim Process is done.

Fig. 8. The user interface for Sustainable Patent Mining before search (top) and after search (bottom)

The results retrieved using the sustainable design indicators are subsequently sorted in a decreasing order of relevance to the query and are displayed in a new pop-up window (only top 20 patents with the highest similarity are listed for each sustainable design indicator) when a sustainable design indicator button is clicked as shown in Fig. 9. This is unlike modern search engines which only focuses on only a single aspect of potential similarities (e.g. semantic relations amongst pieces of texts (words)).


Patent Mining V 1.0.1
Open File
OK
Environmental
high material utilization
energy and resource conservation
biodegradable material
material easy recovery
identifiable material
material of low production
Results (biodegradable material)

Please enter your csv filename (ex: filename.csv):

Patent Data (sustainable)1.csv

The data collection process is done.

The data processing is done.

The Gensim Process is done.

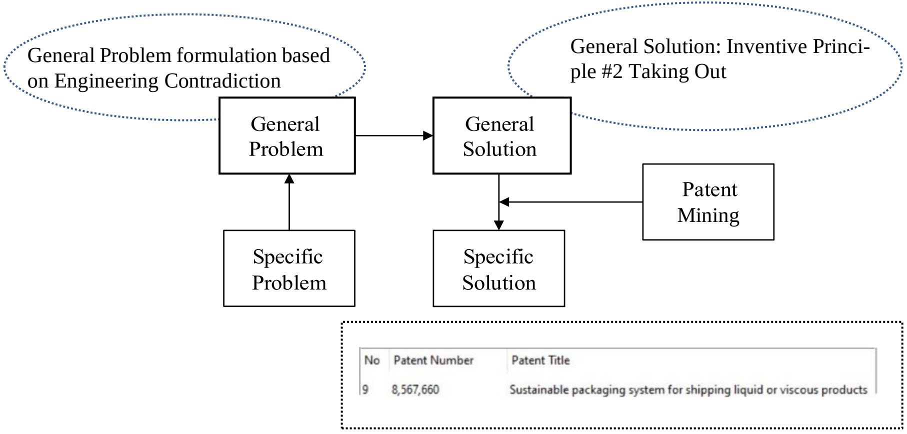
No	Patent Number	Patent Title
0	10,550,010	Oleophilic foams for oil spill mitigation
1	10,543,517	Particle for treatment and repairing of soil contamination and repairing method
2	8,809,843	Nickel-based electrocatalytic photoelectrodes
3	10,577,511	Stretchable hydrophobic materials and methods for making the same
4	9,488,393	Adsorption cell, adsorption cluster and adsorption compressor as well as method of operating an adsorption compressor
5	7,699,896	Surface pretreatment of fabrics for laser writing applications
6	5,979,477	High purity, non-contaminating, burst disk
7	8,980,431	Primer composition for optical articles and optical articles
8	8,919,989	Piezoelectric generator
9	8,567,660	Sustainable packaging system for shipping liquid or viscous products
10	6,406,636	Methods for wafer to wafer bonding using microstructures
11	9,918,451	Livestock feeder
12	9,380,761	Livestock feeder
13	9,918,451	Livestock feeder
14	9,380,761	Livestock feeder
15	9,428,635	Coating of hydroxylated surfaces by gas phase grafting
16	7,255,881	Metal-containing materials
17	10,111,501	Self-binding strap device and method for using same
18	8,905,742	Compact rotary platen 3D printer
19	7,335,076	Swim training fin

Fig. 9. The user interface for Sustainable Patent Mining after the "Biodegradable" button is clicked

4 The Application of Sustainable Product Innovation using Patent Mining and TRIZ

To reiterate a framework which integrates TRIZ, and patent mining aims to support engineers to design sustainable products or solutions. A case study has been conducted to derive a sustainable solution for home liquid detergents (with containers of diverse sizes and made of different materials) which are widely sold in the market. After use, most of these containers are not recycled nor reused and thus, causes environmental issues in the long run due to its non-biodegradable materials (i.e. polymers). Hence, the essential requirement for the design of a liquid detergent container is to facilitate recycle or reuse (duration of action of stationary object feature) to conserve resources. However, the container will not be able to accommodate different detergent volume (adaptability or versatility feature). The inventive principles can be identified using a TRIZ contradiction matrix table and the recommended inventive principle is 'Taking Out' which easily implicates the replacement of the existing container of various sizes and material to a re-usable or recyclable type. However, this leads to the pertinent question on how to design a re-usable and sustainable container that could accommodate different volumes of liquid detergent.

With this inventive principle as a guide, engineers can start looking for specific solution ideas through the patent documents recommended by a patent mining system based on sustainable design indicators shown in Table 1. Engineers can explore the search results of patent mining to elicit design suggestions which are related to the recommended inventive principle. For example, in the results of “Bio-degradable material”, the design suggestion with patent number 8567660 is a patent that describes a “multi-layer packaging design that can be used to contain liquid” which can be adapted and applied to replace the existing container that concurs with the inventive principle #2 (Taking Out) which is shown in Fig. 10. Therefore, the designer can study the details of the patent which will assist engineers derive a specific solution to solve the liquid detergent container problem. This demonstrates how patent mining and TRIZ can be combined to help engineers to find solutions to the problem.


General Problem formulation based on Engineering Contradiction
General Solution: Inventive Principle #2 Taking Out
General Problem
General Solution
Patent Mining
Specific Problem
Specific Solution

No	Patent Number	Patent Title
9	8,567,660	Sustainable packaging system for shipping liquid or viscous products

Fig. 10. Suggestion which relates to Inventive Principle 2 (Taking Out) assisted with a result from patent mining system.

5 Results and Discussions

Fig. 9 shows the results of the potential suggestions for sustainable design with the 20 top similarity results of patent documents to the sustainable design indicator “Biodegradable material”, patent number, and patent title. The listing of these top similarity results of patent documents are based on their score in similarity calculated by Cosine measure (-1 to 1) where the similarity result of 1.0 is highest in regard to the sentences in the patent titles and abstracts. The search results only show the top 20 results sorted in descending order to reduce computational time. Engineers can also explore other sustainable design indicators to explore the details of potential solutions for finding specific solutions to replace the liquid detergent container. To reiterate, with the ideas from specific solution of patent 8567660 as shown in Fig. 11, a novel re-usable container can be designed with a rigid recyclable material such as kenaf as the outer container with an internal detachable bag made of biodegradable plastics that is waterproof of different sizes to cater for varying volume of liquid detergent.

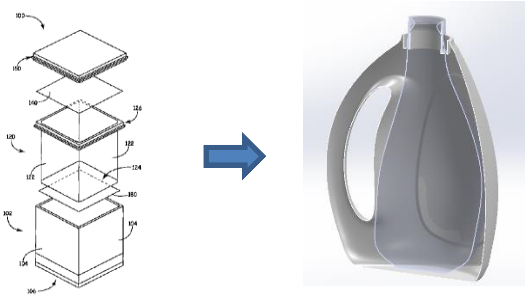


Fig. 11. Results of specific solution from USPTO Patent No. 8567660 (left) that inspire the sustainable design (right)

6 CONCLUSION AND FUTURE WORK

In this paper, a methodology of patent mining that used to analyze patent documents for TRIZ users is presented. Firstly, the patent document is retrieved from USPTO with HTML format and imported into Python workspace. Next the database is pre-processed to segment the lengthy database into tokens and the stopwords are filtered out. Subsequently, every individual word is normalized into its base form so that the size of the dataset can be reduced. The patent mining results are discussed and shown with an example of how patent mining result can be linked with the inventive principle of TRIZ.

For future work, the current algorithms could be further improved for enhancing the performance of document classification algorithms. Claims or descriptions can be used as the corpus because the abstracts contain less technical detail information.

Acknowledgements. The authors would like to thank the Universiti Kebangsaan Malaysia and the Ministry of Higher Education, Malaysia, for supporting the work through research grants: GUP-2018-124 and FRGS/1/2018/TK03/UKM/02/6.

REFERENCES

1. Sherwin, C.: Design and sustainability: A discussion paper based on personal experience and observations. The Journal of Sustainable Product Design 4, (2004)
2. Zainali, N.S., Ang, M.C., Ng, K.W., Ijab, M.: A Framework for Sustainable Eco-Friendly Product Development Based on TRIZ. In: al., H.B.Z.e. (ed.) Advances in Visual Informatics, vol. 11870, pp. 704-712 (2019)
3. Kor, A.L., Rondeau, E., Andersson, K. (eds.): AI for Sustainability and Innovation [Special Issue], Vol. 11. Applied Sciences (2021)
4. UN: The 17 Goals. Department of Economic and Social Affairs (2021)

5. Cao, G., Luo, P., Wang, L., Yang, X.: Key Technologies for Sustainable Design Based on Patent Knowledge Mining. Procedia CIRP 39, 97-102 (2016)

6. Waas, T., Hugé, J., Verbruggen, A., Wright, T.: Sustainable Development: A Bird's Eye View. Sustainability 3, 1637-1661 (2011)

7. Ang, M.C., Ng, K.W., Ahmad, S.A., Wahab, A.N.A.: An Engineering Design Support Tool Based on TRIZ. In: al., H.B.Z.e. (ed.) 3rd International Visual Informatics Conference (IVIC 2013), pp. 115-127. Springer International Publishing, Switzerland. Lecture Notes in Computer Science 8237, Equatorial Hotel Bangi, Selangor, Malaysia (2013)

8. Hua, Z., Yang, J., Coulibaly, S., Zhang, B.: Integration TRIZ with problem-solving tools: A literature review from 1995 to 2006. International Journal of Business Innovation and Research - Int J Bus Innovat Res 1, (2006)

9. Li, M., Ming, X., Zheng, M., He, L., Xu, Z.: An integrated TRIZ approach for technological process and product innovation. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture 231, 1062-1077 (2015)

10.Fischer, G., Lalyre, N.: Analysis and visualisation with host-based software-The features of STN®AnaVist^{TM}. World Patent Information 28, 312-318 (2006)

11.Soo, V.-W., Lin, S.-Y., Yang, S.-Y., Lin, S.-N., Cheng, S.-L.: A cooperative multi-agent platform for invention based on patent document analysis and ontology. Expert Systems with Applications 31, 766-775 (2006)

12.Tseng, Y.-H., Lin, C.-J., Lin, Y.-I.: Text mining techniques for patent analysis. Information Processing & Management 43, 1216-1247 (2007)

13.Yoon, B., Park, Y.: A text-mining-based patent network: Analytical tool for high-technology trend. The Journal of High Technology Management Research 15, 37-50 (2004)

14.Ghane, M., Ang, M.C., Kadir, R.A., Ng, K.W.: Technology Forecasting Model Based on Trends of Engineering System Evolution (TESE) and Big Data for 4IR. 2020 IEEE Student Conference on Research and Development (SCOReD), pp. 237-242 (2020)

15.Liang, Y., Tan, R.: A Text-Mining-based Patent Analysis in Product Innovative Process. In: Trends in Computer Aided Innovation, pp. 89-96. Springer US, (Year)

16.Yusof, S.M., Awad, A.A.: A Brief Review of Theory of Inventive Problem Solving (TRIZ) Methodology. Jurnal Teknik Industri – Universitas Bung Hatta 2, 119-131 (2014)

17.Bird, S.: NLTK: The natural language toolkit (2006)

18.Ballard, N., Joshi, D.: Similarity matching in news articles. J. Comput. Sci. Coll. 35, 46–51 (2019)

19.Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques. Morgan Kaufmann, Boston (2012)