A Conceptual Design Framework based on TRIZ Scientific Effects and Patent Mining

E-Ming Chan[1]
Department of Mechanical, Materials and Manufacturing Engineering
University of Nottingham Malaysia, 43500 Semenyih, Selangor, Malaysia

Ah-Lian Kor[2]
School of Built Environment, Engineering and Computing
Leeds Beckett University, Leeds, United Kingdom

Kok Weng Ng[3]
Department of Mechanical, Materials and Manufacturing
Engineering, University of Nottingham Malaysia, 43500
Semenyih, Selangor, Malaysia

Mei Choo Ang[4]*
Institute of IR4.0, Universiti Kebangsaan Malaysia
43600 UKM, Bangi, Selangor, Malaysia

Amelia Natasya Abdul Wahab[5]
Faculty of Information Science and Technology
Universiti Kebangsaan Malaysia
43600 UKM, Bangi, Selangor, Malaysia

Abstract—Conceptual design represents a critical initial design stage that involves both technical and creative thinking to develop and derive concept solutions to meet design requirements. TRIZ Scientific Effects (TRIZSE) is one of the TRIZ tools that utilize a database on functional, transformation, parameterization of scientific effects to provide conceptual solutions to engineering and design problems. Although TRIZSE has been introduced to help engineers solve design problems in the conceptual design phase, the current TRIZSE database presents general scientific concept solutions with a few examples of solutions from patents which are very abstract and not updated since its introduction. This research work explores the derivation of a novel framework that integrates TRIZ scientific effects to the current patent information (USPTO) using data mining techniques to develop a better design support tool to assist engineers in deriving innovative design concept solutions. This novel framework will provide better, updated, relevant and specific examples of conceptual design ideas from patents to engineers. The research used Python as the base programming platform to develop a conceptual design software prototype based on this new framework where both the TRIZSE Database and Patents Database (USPTO) are searched and processed in order to build a Doc2Vec similarity model. A case study on the corrosion of copper pipelines by seawater is presented to validate this novel framework and results of the novel TRIZSE Database and patents examples are presented and further discussed in this paper. The results of the case study indicated that the Doc2Vec model is able to perform its intended similarity queries. The patent examples from results of the case study warrant further consideration in conceptual design activities.

Keywords—*TRIZ; patent mining; natural language processing; product design*

I. INTRODUCTION

TRIZ is a Russian Acronym which translate to the "Theory of Inventive Problem Solving" which was first initiated by Genrich Altshuller in 1946 [1]. TRIZ represents an engineering theory consisting of many problem-solving tools derived from compiling and analyzing utility patents to find general solutions and trends, which will be used to solve complex engineering problems that can be further decomposed into their individual problems and components.

One of the main TRIZ tools is the contradiction matrix which is well-established and is used to improve an engineering system based on solving engineering contradictions. According to TRIZ, there will always be engineering contradictions that need to be resolved in all inventive problems [2]. Contradiction matrix will recommend inventive principles derived from years of study of patent information to assist engineers resolve these contradictions. However, these inventive principles are very general and abstract for most engineers and interpreting these inventive principles in the context of specific solution concepts are challenging because expertise knowledge and experience are required. The other TRIZ tools including TRIZ scientific effects (TRIZSE) and system of standard inventive solutions also recommend solution concepts which are very general and abstract [1]. Therefore, research have explored into extending and improving existing TRIZ tools and proposed new patent extraction system such as Inventive Design Method (IDM) and extracting design information from patent documents and focus on assisting engineers to overcome their psychological barrier [3]. The proposed IDM has its limitations and deficiencies in issues related to duplicated information. This research work acknowledges the importance and criticality of defining design problems described in the form of function, parameter, transform and a real-time link to patent documents in patent offices to provide accurate relevant examples to assist engineers.

Therefore, this research work will explore the derivation of a framework to link TRIZSE to patent documents to support engineers in conceptual design and investigate the potential of enhancing TRIZSE with patents database. By developing this

---

*Corresponding Author.

link, TRIZSE may provide a better and clearer solution to each of their effects. The main objectives of this paper are:

1) To introduce the preliminary conceptual design system using TRIZSE which is enhanced by utilizing a similarity model built on a patent database to provide sufficient examples suggested by TRIZSE.

2) To build a Graphical User Interface (GUI) to allow a user to provide their input and queries for a single functional statement.

3) Conduct a case study to assess the performance of the program in factors of user input, data accuracy and its relevance.

The rest of the paper will cover the extraction of information, development of Doc2Vec model, and a case study. The proposed process model will be presented in the methodology section. This is followed by validation of the Doc2Vec model for feature extraction from the patent database for a use case. The report will then be concluded with recommendation for future improvement.

II. BACKGROUND

TRIZ Scientific Effects (TRIZSE) is one of the TRIZ tools which is applied in the conceptual design phase that utilizes a database based on three types of search query; Functional, Parameterization, and Transformation to recommend a list of solution concepts (for e.g., if a user desires to constrain gas, one suggestion effect from the Functional aspect is Glassy Carbon which has high temperature resistance, extreme resistance to chemical attack, and impermeability to gases and liquid [4]). However, these suggestions lack suitable examples to support the recommended list of solution concepts and fail to inspire conceptual design ideas. Nonetheless, these deficiencies of the TRIZSE database offer a unique opportunity to explore the possibility of applying appropriate data mining techniques to search for examples related to the recommended list of solution concepts based on available online patents information with the intention of retrieving relevant patents as examples for any given recommended solution concepts by TRIZSE. To link the recommended solution concepts of TRIZSE to patent information, it is necessary to develop a suitable framework (with a user-friendly user interface) that integrates data analysis and mining technique (also known as patent mining) based on TRIZSE.

Using similar initiative to enhance a TRIZ technique, TRIZ Scientific Effects (TRIZSE) are comprehensive descriptions of effects based on sciences which provide good conceptual solutions which can potentially link patent information to obtain sufficient relevant real examples from TRIZSE database. In conceptual design activities, TRIZSE plays a role of redefining each function in a system and provides a set of suggestions of functional solutions as shown in an online database developed by Martin [4]. TRIZSE itself serves as a basis of creative idea generation in solving an engineering problem. Access to TRIZSE databases is limited due to subscriptions or memberships [5, 6], and third party state-of-the-art patent mining and analysis software [7]. However, the TRIZSE database provided by Martin [4] is available for public usage.

For the conceptual design framework development, the user must first define the engineering problem to be solved. The user-defined problem should be decomposed from complex problems into simple sentence structure before being used as an input. This not only simplifies the problem but also provides a clearer view of expected results from the query. There are vast selections of problem analysis tools such as Causal Loop diagram [8] that identify the polarity influences, Functional Decomposition and Morphology [9] decompose complex systems and re-merges as well as Systems Dynamics [10] that can simulate potential design changes. However, a simple and direct tool should be selected as it can provide more direct input for the model to manipulate. Hence, Functional Analysis (FA) [11] which maps component relationships can be selected for deployment. FA fosters problem statement formulation that could be further decomposed into a simple solution statement consisting of Subject-Action-Object. This facilitates better transition to the parameters of the TRIZSE database.

The linkage between TRIZSE to a patent database can be implemented through text mining techniques. A text mining Python library called Gensim is widely popular amongst patent mining research and has diverse range of data mining models. One of the models is a similarity model Doc2Vec inspired by vector representations of words using neural networks known as Word2Vec [12]. The Word2Vec model has a disadvantage as it does not consider the word order of each text document [12]. It functions similarly to a bag-of-words model where only the frequency of each word in the patent document is accounted for [12]. The proposed Doc2Vec is an improved model that can assess the word and neighboring words in the sentence to properly gauge the semantics of the entire sentence. Doc2Vec is compared to other similarity models such as Term frequency-inverse document frequency (TF-IDF) which calculates the importance of a word based on the frequency in the patent document and Global Vectors for Word Representation (GloVe) just like TF-IDF but with word embeddings to discover word contexts [13]. The result of comparisons shows that Doc2Vec is superior and capable of handling higher amounts of data entries (in millions) [13].

The patents database is sourced online. There are major databases such as the United States Patents and Trademark Office (USPTO), the Japanese Patent Office (JPO), the European Patent Office (EPO) [14]. For the ease of database extraction, the USPTO database is selected.

III. METHODOLOGY

A. Data Extraction

The TRIZSE database is extracted from the online database [4]. A third party free-to-use software called ParseHub is shown in Fig. 1. It performs as an automated user-action and data collector. Parsehub is utilised to extract every combination of the database and its results. The data extracted is stored into a Comma-Separated Values (csv) file.

Patent documents are unstructured texts with sections of irregular word limits and formatting. Thus, this paper makes use of one of the consistent sections of a patent document, the abstract. Sections such as claims would range from few sentences to several pages.
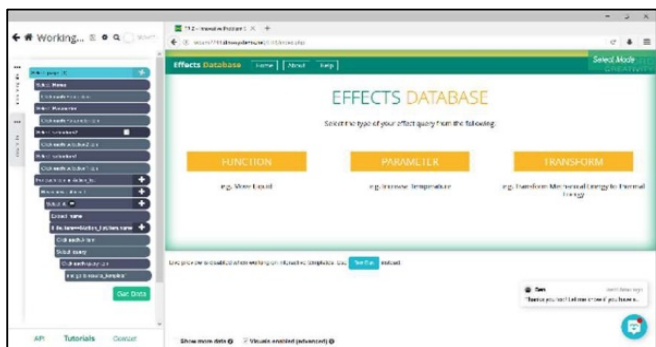
Fig. 1. Brief Interface of Parsehub.

B. Data Processing

The main data processing involves handling of millions of patents in the database. In USPTO, there are multiple types of patents such as utility, design, plant. Utility patents are patents that contain a feature or solution whereas design patents are typically drawings. Only utility patents are added value documents for this study. The total number of utility patents from the year 1971 to 2020 numbers is approximately 6.8 million. The abstracts of patents are useful for feature extraction in patent mining as there is a word limit and crucial information is present. Hence, each patent document's abstract undergoes a pre-processing stage, where words will be separated, and reformatted. The pre-processing steps involve removal of special characters, lowercasing capitals, and removal of redundant words [15, 16]. The pre-processing technique is adapted from the Gensim code examples [17]. The patent documents are tagged by simple indexing for accessing the documents in later stages.

C. Similarity Model Technique and Concept

Similarity models encompass supervised and unsupervised techniques, each with their own advantages and disadvantages. Supervised techniques involve high percentage of manual sorting and tagging that yields better overall results. On the other hand, unsupervised techniques do not have manual inspection. However, it is preferable over supervised techniques when dealing with massive number of documents. The Doc2Vec model is an unsupervised technique model. The data is fed into the algorithm and all unique words will be extracted from the patent documents into a vocabulary dictionary. The Doc2Vec model starts training with a vocabulary size and input texts. The Doc2Vec modelling can be simplified (as shown in Fig. 2).


```
Doc2Vec(abstracts, min_count=2,
window=3, vector_size=100,
epochs=20, workers=7)
```

Fig. 2. Modelling of Doc2Vec.

The description of the arguments for Doc2Vec is as follows: abstracts represent the input of patent abstracts; min_count the removal of words with frequencies less that the value f (in this example, 2); window is N (3 in this example),

number of neighboring words that can affect the word of interest; vector_size shows the number of numerical elements that an abstract vector can attain. The higher the number of vector size, the higher the number of features of an abstract that can be captured; epochs are the number of training iterations of the model; workers are the number of computer cores for processing. With the vector generated, it can be compared numerically with a cosine similarity approach [14][17].

D. Process Chart of the Conceptual Design Framework

Once trained, the model can produce and predict a vector for new inputs to reflect similarity results. This can be used to evaluate similarity of a TRIZ Scientific Effect concept with real patent examples. The development of a Graphical User Interface (GUI) using PySimpleGUI library is to facilitate user input. The activities are depicted in Fig. 3. The processes and output is visible to the user through Graphical User Interface. The user inputs their problem in a Functional statement which will be a guide to invoke suitable TRIZSE settings. TRIZSE database will return results from the inputs which are fed into the similarity model individually by each effect. The model results are compiled in a descending order of patent similarity.
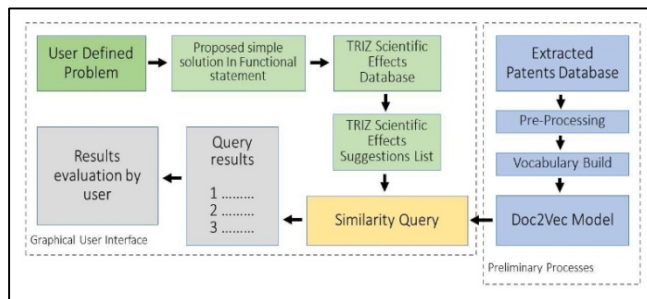

Fig. 3. Process Chart of the Conceptual Design System Framework.

IV. RESULTS AND DISCUSSION

A. TRIZSE Database Extraction

This work starts with extraction and analysis of the TRIZSE database. The database extracted from the source website takes approximately 15 hours to parse through all combinations. The CSV file saved is checked and validated for accuracy. Fig. 4 shows the sample of the data extracted from the database. Each combination of TRIZ Scientific Effect yields a set of results which are suggestions of Effects along with brief description of the concept.

Absorb	Divided Solid	Effect	17 suggest Amphiphiles	A chemical compound poss
Absorb	Divided Solid	Effect	17 suggest Bingham Plastic	A viscoplastic material that
Absorb	Divided Solid	Effect	17 suggest Diffusion	The movement of particles
Absorb	Divided Solid	Effect	17 suggest Electret	An Electret is a dielectric m
Absorb	Divided Solid	Effect	17 suggest Fractal Forms	A fractal is generally 'a roug
Absorb	Divided Solid	Effect	17 suggest Gel	A gel is a solid, jelly-like ma
Absorb	Divided Solid	Effect	17 suggest Ostwald Ripening	An observed phenomenon
Absorb	Divided Solid	Effect	17 suggest Oxidation	A chemical reaction that in

Fig. 4. A Snapshot of Extracted TRIZSE Database.

The data extracted undergoes data elimination of duplicated data. For example, a combination of Absorb and Divided Solid in Function section can have three types of results which are

Framework, Effect, or Both. The results that are repetitive can be merged. For example, the Functional section containing 21,175 rows of effect suggestions is merged into 10,589 rows, which is 50% smaller. Similar process is applied to Parameter and Transformation section which yields similar reduction.

The Effects in TRIZSE database are obtained through the various combinations of inputs. Each Effects accommodates multiple combinations, and the total of Effects is approximately 1000. The concept description of each Effects will be assessed by the Doc2Vec model to identify similar concepts in patents.

B. USPTO Patents Extraction

The USPTO patents are downloaded from PatentsView website in a compressed ZIP file. A Python code parses through the ZIP file to extract the desired information which comprises the following: utility patents' number, title, and abstract. The process has taken 5 hours for 7.5 million patent documents and 6.8 million utility patents are extracted and stored.

C. Doc2Vec Model Code Development

The Doc2Vec modelling follows the stages of Preliminary processes in Fig. 3. The first stage is the pre-processing of each patent abstracts. In this stage, abstracts will undergo tokenization. Tokenization splits the abstracts, a paragraph of words, into individual words while keeping the order of the sentence. Each individual word is converted to lowercase and stop words are removed. Stop words are words that do not add to sentence context such as "a", "of", "and", "also". The duration of the pre-processing stage is 2660 seconds. Prior to Doc2Vec modelling, the abstracts undergo a vocabulary building phase which identifies unique words and stores it into a dictionary. This phase identifies the weight of each word. Heavier word weight shows higher importance. The process is 4340 seconds. The Doc2vec modelling uses the settings from Fig. 2, frequency of words less than 2 will be removed since they are uncommon words. The window size of 3 influences the target word with 3 nearby words. The vector_size is also known as feature size that captures the feature of a document and converts it into a vector of numbers [12] that is calculated from the weights of words in the vocabulary dictionary. The vector_size is set to 100 which balances between memory usage and feature uniqueness. Epochs are set at 20 with each iteration taking 40 minutes. By running multiple iterations on the data, the Doc2Vec model learns and re-calibrates features in a document. The Doc2Vec model is successfully built into 6.8 million abstract vectors in a process that has taken 17 hours. The model loaded within the Python workspace is displayed in Fig. 5 showing it has been successfully booted.

The unique tokens and the number of documents vectorized in the model is shown in Fig. 6.

D. Doc2Vec Model Validation

The model can be validated by a similarity query of a random patent abstract section from USPTO. Typically, the result of similarity in the model will proceed with high similarity documents, followed by lower similarity documents. The output in Fig. 7 shows that the behavior is as predicted.

The output seen in Fig. 7 uses a section of the abstract as input to find similar documents. The first result has a similarity of 88.57% and on online inspection, found to be the correct patent. Thus, the model can provide accurate similarity. A similarity tolerance test is conducted by running 10 times on the same query shown in Fig. 8. The graph shows that the average similarity value is at 0.885825 with a range of ± 0.000144. Since the value of the range is significantly small, it will not affect the order of top documents in similarity results.

Fig. 5. Doc2Vec Model Logs.

Fig. 6. Output of Token and Vectors Computed.

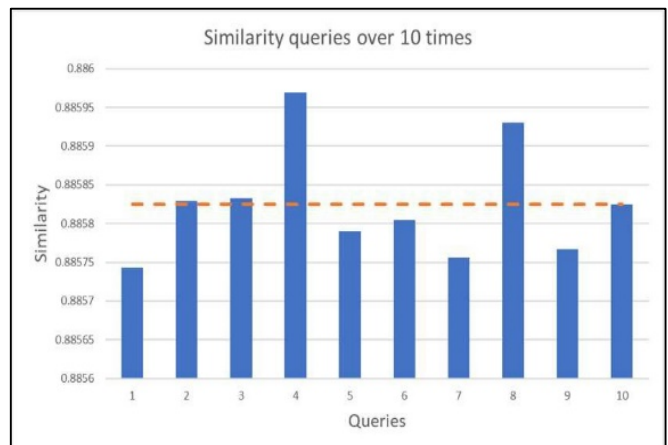Fig. 7. Similarity Query of a Patent Abstract.

Fig. 8. Plotted Percentages of Similarity Query.

E. Graphical user Interface (GUI) Development

Once the fundamental operation of the program is developed, a GUI is built to accommodate user input. The interface contains the user input of their Functional Statement in three text boxes: Subject, Action, Object. With the guidance of the proposed Functional statement simple solution input, the rest of the TRIZSE database can be selected. The TRIZSE database is laid out with Function, Parameter and Transform sections clearly indicated. The user may select the applicable selections from the available options as well as the ability to exclude un-used sections. The GUI is coded and loaded as shown in Fig. 9, provides enough information to users to perform their input activities.

Fig. 9. Graphical user Interface of the Conceptual Design Software Prototype.

F. Case Study on the Corrosion of Copper Pipelines caused by Seawater

Copper alloys are used in ships as pipelines, heat exchanges, screw propellers, valves [18] and most components are often exposed to seawater which is corrosive. In marine engineering applications, copper alloys have an innate seawater corrosion resistance, but will be vulnerable under heavier load conditions [18] such as pressure generated from propeller torque or oscillating temperature in pipeline cooling systems. The case study aims to discover methods to prevent the corrosion of copper pipelines in a marine boat's cooling system. The problem first is approached by using Functional Analysis to identify and break down the problem into a simple problem statement of Subject-Action-Object as shown in Fig. 10.
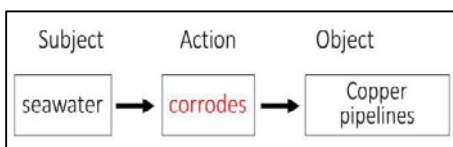

Fig. 10. Problem Statement in Functional Analysis Form.

The simplified problem statement is transformed into a simple solution statement by reversing the subject, object, and providing a counteraction. The solution statement is presented using the same format as shown in Fig. 11.
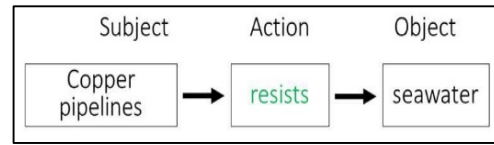

Fig. 11. Solution Statement in Functional Analysis Form.

With the defined solution statement, the GUI can be filled together with the TRIZSE options. Corrosion is an electrochemical reaction that requires surface contact and seawater is a form of liquid. Thus, in the Functional section, the best combination will be to resist liquid as an application. To prevent contact between copper and seawater, the surface finish should not degrade and should have high wear resistance. This can be used in the parameterization section with stabilize surface finish as an application. In the final section, it is unnecessary as the solution should be a preventive measure and thus, assumed to have no transformation of energy in copper pipelines. With the information gathered, the software prototype can be launched and have the input arranged in the GUI as displayed in Fig. 12. From the input screen, it is then submitted and the results of the TRIZSE database are shown in Fig. 13 and Fig. 14 which will be used with the Doc2Vec model to discover similar patent examples.
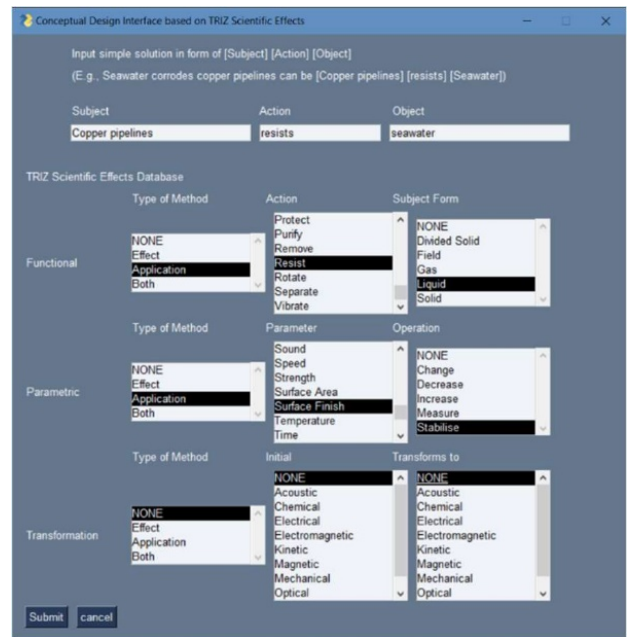
Fig. 12. GUI Interface with Case Study Settings.

As observed in Fig. 13, the list of suggestions from TRIZSE database provides a brief description of each effect with limited examples. A brief review will be conducted to reduce the redundant effects generated. Effects such as Valve and Tesla Valvular Conduit can be removed from consideration as it is an interest of internal volume. The effect Intumescent materials will also be removed as temperature should not be a factor in corrosion resistance. Another similar

review will be conducted on the results of the Parametrization section shown in Fig. 14. From the results of the Parametrization section, some effects can be removed to refine the patent results. Effects such as Redundancy may be considered as a last resort stage and will be removed. The effect Thin Films is also present in the list in Fig. 13 and will be removed since both lists will be compiled before loading into the Doc2Vec model. Using each TRIZSE obtained from the input, it is used together with the subject as a query input to the Doc2Vec model.

Input		TRIZ Scientific Effects	Effect Descriptions
Resist	Liquid	Diamond-like Carbon	(DLC) exists in seven different forms of amorphous carbon materials that display some of the unique properties of natural diamond. They are usually applied as coatings to other materials. All seven contain significant amounts of sp3 hybridized carbon atoms. As well as excellent hardness and wear resistance, DLC is claimed to have the lowest coefficient of friction of any known solid material.
Resist	Liquid	Ferrofluid	A liquid which becomes strongly magnetized in the presence of a magnetic field. Ferrofluids are colloidal mixtures composed of nanoscale ferromagnetic, or ferrimagnetic, particles suspended in a carrier fluid, usually an organic solvent or water. The ferromagnetic nanoparticles are coated with a surfactant to prevent their agglomeration (due to van der Waals and magnetic forces).
Resist	Liquid	Intumescent Materials	An intumescent is a substance which swells as a result of heat exposure, thus increasing in volume, and decreasing in density. Intumescent are typically used in passive fire protection.
Resist	Liquid	Polytetrafluoroethylene (PTFE)	A synthetic fluoropolymer. As fluorocarbons are not as susceptible to van der Waals force (due to the high electronegativity of fluorine) water and water-containing substances, and oil and oil-containing substances, like most foods do not wet PTFE. PTFE is very non-reactive. PTFE's coefficient of friction is 0.1 or less, which is the second lowest of any known solid material.
Resist	Liquid	Solenoid	A coil wound into a tightly packed helix. In physics, the term solenoid refers to a long, thin loop of wire, often wrapped around a metallic core, which produces a magnetic field when an electric current is passed through it. Solenoids are important because they can create controlled magnetic fields and can be used as electromagnets.
Resist	Liquid	Tesla Valvular Conduit	A one-way valve with no moving parts, using the geometry of the fluid path to redirect the flow of the fluid to oppose its own motion in one direction, but offering little resistance in the other.
Resist	Liquid	Thin Films	Thin material layers ranging from fractions of a nanometer to several micrometers in thickness. Electronic semiconductor devices and optical coatings are the main applications benefiting from thin film construction.
Resist	Liquid	Valve	A device that regulates the flow of a fluid (gases, liquids, fluidized solids, or slurries) by opening, closing, or partially obstructing various passageways.

Fig. 13. Output of Functional Section Effects Database.

Input settings		TRIZ Scientific Effects	Effect Descriptions
Stabilize	Surface Finish	Coatings	A covering that is applied to an object. The aim of applying coatings is to improve surface properties of a bulk material usually referred to as a substrate. One can improve amongst others appearance, adhesion, wettability, corrosion resistance, wear resistance, scratch resistance, etc. They may be applied as liquids, gases or solids.
Stabilize	Surface Finish	Electroplating	The process of using electrical current to reduce cations of a desired material from a solution and coat a conductive object with a thin layer of the material, such as a metal. Primarily used to bestow a desired property (e.g., abrasion and wear resistance, corrosion protection, lubricity, aesthetic qualities, etc.) to a surface that otherwise lacks that property. Also used to build up thickness on undersized parts.
Stabilize	Surface Finish	Epitaxy	The method of depositing a monocrystalline film on a monocrystalline substrate. Epitaxial films may be grown from gaseous or liquid precursors. Because the substrate acts as a seed crystal, the deposited film takes on a lattice structure and orientation identical to those of the substrate.
Stabilize	Surface Finish	Feedback	A circular causal process whereby some proportion of a system's output is returned (fed back) to the input. This is often used to control the dynamic behavior of the system.
Stabilize	Surface Finish	Ferromagnetic Powder	Ferromagnetic material in a powdered or finely divided form. Ferromagnetic materials (such as iron) form permanent magnets and/or exhibit strong interactions with magnets. Ferromagnetic materials lose their ferromagnetic properties above a characteristic temperature (the Curie Point).
Stabilize	Surface Finish	Physical Vapor Deposition	A variety of methods used to deposit thin films by the condensation of a vaporized form of the material onto various surfaces
Stabilize	Surface Finish	Preservative	A naturally occurring or synthetic substance that is added to products such as foods, pharmaceuticals, paints, biological samples, wood, etc. to prevent decomposition by microbial growth or by undesirable chemical changes.
Stabilize	Surface Finish	Redundancy	The duplication of critical components of a system with the intention of increasing reliability of the system, usually in the case of a backup or fail-safe.
Stabilize	Surface Finish	Sintering	A method for making objects from powder, by heating the material below its melting point (solid state sintering) until its particles adhere to each other. Traditionally used for manufacturing ceramic objects. Most, if not all, metals can be sintered - especially pure metals produced in vacuum which suffer no surface contamination. Many nonmetallic substances also sinter, such as glass, alumina, zirconia, silica, magnesia, lime, ice, beryllium oxide, ferric oxide, and various organic polymers.
Stabilize	Surface Finish	Thin Films	Thin material layers ranging from fractions of a nanometer to several micrometers in thickness. Electronic semiconductor devices and optical coatings are the main applications benefiting from thin film construction.

Fig. 14. Output of the Parametrization Section Effect Database.

Each query has been set to have an output of the first 10 patents and this setting can be customized as a preference. However, since the similarity of documents is highly saturated in the first few results, the similarity of the rest of the documents may be too broad to provide substantial design support. The output results are compiled, and the top 5 results are shown in Fig. 15. From displayed results in Fig. 15, the outcome of the top 5 highest similarity in patent documents have the effect Physical Vapor Deposition in majority. Analysis of the first patent (10115603) reveals that the patent describes a method of removing passivation films. Passivation film is a coat of protective material typically able to protect against corrosion [19]. However, the patent suggests removal of the said film instead. This displays the model inaccuracy of polarity checking, although being accurate to the topic of corrosion. The second patent (5514414) discusses the deposition of heated flux vapor by condensing onto the subject.

The role of flux vapor in corrosion resistance is by removing impurities and oxidation layer [20]. The third patent (9859218) is not applicable towards pipelines as its application is focused on reinforcement of interconnected structures. The fourth patent (5352331) focuses on electroplating a thin film cermet; a composite of ceramic (cer) and metal (met). A thin layer of cermet is capable of producing corrosion resistance and protecting the internal material [21]. The fifth patent (4927472) discusses the formation of tin phosphate on metal surface which creates a layer of corrosion resistance. The idea of phosphating a base material with solution of phosphate ions can provide for corrosion resistance against highly corrosive environments such as seawater [22]. From the first 5 results of the Doc2Vec model, although not all results have the intention of offering corrosion resistance, it is still relevant to the topic scope of the model input.

Effect	Similarity Number	Patent	Title	Abstract
Physical Vapor Deposition	63.45%	10115603	Removal of surface passivation	Methods for removing a passivation film from a copper surface can include exposing the passivation film to a vapor phase organic reactant for example at a temperature of 100° C. to 400° C. In some embodiments the passivation film may have been formed by exposure of the copper surface to benzotriazole such as can occur during a chemical mechanical planarization process. The methods can be performed as part of a process for integrated circuit fabrication. A second material can be selectively deposited on the cleaned copper surface relative to another surface of the substrate.
Physical Vapor Deposition	61.86%	5514414	Solvent-less vapor deposition apparatus and process for application of soldering fluxes	Apparatus and method for condensing a solderless flux vapor onto a work surface to be soldered such as an electronic circuit board. The flux vapor is created by heating flux in a liquid state to a temperature greater than the temperature of the work surface. Flux is applied to the work surface without the use of any volatile organic chemicals.
Physical Vapor Deposition	58.86%	9859218	Selective surface modification of interconnect structures	Semiconductor structures including copper interconnect structures and methods include selective surface modification of copper by providing a CuxTiyNz alloy in the surface. The methods generally include forming a titanium nitride layer on an exposed copper surface followed by annealing to form the CuxTiyNz alloy in the exposed copper surface. Subsequently the titanium layer is removed by a selective wet etching.
Electroplating	58.29%	5352331	Cermet etch technique for integrated circuits	An etching process for patterning thin film cermet (14) on a semiconductor substrate (10) using a mild room temperature acid solution as the etchant. The semiconductor substrate (10) has a glass passivating layer (12) such as silicon dioxide deposited thereon. The cermet layer (14) is deposited on the silicon dioxide layer (12). A photoresist layer (16) is deposited and patterned on the cermet layer (14) followed by the deposition of a layer of aluminum (18). The cermet (14) is then preferentially etched with a mild room temperature hydrofluoric acid solution diluted with hydrochloric acid to form the desired cermet resistance pattern.
Physical Vapor Deposition	56.67%	4927472	Conversion coating solution for treating metal surfaces	A conversion coating predominantly consists of tin phosphate can be deposited on steel or tin-plated steel surfaces by contact with a solution containing phosphate ions tin ions an oxidizing agent such as chlorate and a chelating agent for the tin ions the latter to prevent the rapid loss of tin from the solution that otherwise would occur. A coating that confers excellent resistance to corrosion in hot water is formed on drawn and ironed thinly tin-plated cans.

Fig. 15. Output of the Doc2Vec Model.

V. CONCLUSION

This research has demonstrated the potential of linking TRIZSE to USPTO patents pool to provide additional support for conceptual design related activities. With the successful building of the Doc2Vec model, a case study has been conducted on the corrosion of copper pipelines in seawater. The concluded discussion of the results of TRIZSE and Doc2Vec model shows that:

- The model provides a majority of relevant examples that ought to be further investigated for the case study.

- The Doc2Vec model does not account for semantics and thus, could not differentiate polarity of relevant examples.

- The improvement of Doc2Vec model can be conducted in stages of pre-processing, feature size, and semantic analysis.

Although the results of patent relevancy are not sufficiently accurate, the linkage of TRIZSE database with USPTO patent pool shows the potential of assisting conceptual design activities. In conclusion, TRIZSE when combined with patent mining provides a viable support means to engineers in developing conceptual solutions by providing relevant examples of solutions in patent documents. Future explorations of the similarity model may be able to provide better accuracy and feature extraction along with semantic analysis. Enhanced accuracy is

necessary as the current search results from the similarity model include accurate relevant concept solutions as well as somewhat inaccurate concept solutions. Future work should only yield search outputs that are only directly relevant patents.

REFERENCES

[1] K. Gadd and C. Goddard, TRIZ for engineers: enabling inventive problem solving. Chichester, West Sussex: Wiley, 2011.

[2] G. S. Altshuller, And suddenly the inventor appeared: TRIZ, the Theory of Inventive Problem Solving. Worcester, Massachusetts: Technical Innovation Center, Inc., 1996.

[3] A. Souili, D. Cavallucci, and F. Rousselot, "Identifying and reformulating knowledge items to fit with the Inventive Design Method (IDM) model for a semantically-based patent mining," Procedia Engineering, vol. 131, pp. 1130 – 1139, December 2015.

[4] A. Martin. "TRIZ Effects database." http://wbam2244.dns-systems.net/EDB/ (accessed 16 December, 2020).

[5] Creax. "Innovation is crucial - secure your company's future." https://creax.com/innovation-services/ (accessed 16 December, 2020).

[6] Samsung. "□□□ □□: TRIZ school." https://www.seri.org/forum/trizschool/ (accessed 16 December, 2020).

[7] Aulive. "Patent inspiration." https://www.patentinspiration.com/ (accessed 16 December 2020).

[8] J. Delgado-Maciel, G. Cortés-Robles, G. Alor-Hernández, J. García-Alcaraz, and S. Negny, "A comparison between the functional analysis and the causal-loop diagram to model inventive problems," Procedia CIRP, vol. 70, pp. 259-264, January 2018.

[9] L. Fiorineschi, F. S. Frillici, F. Rotini, and M. Tomassini, "Exploiting TRIZ tools for enhancing systematic conceptual design activities," Journal of Engineering Design, vol. 29, no. 6, pp. 259-290, June 2018.

[10] J. Delgado-Maciel, G. Robles, C. Sanchez-Ramirez, J. García-Alcaraz, and J. Méndez-Contreras, "The evaluation of conceptual design through dynamic simulation: a proposal based on TRIZ and system dynamics," Computers & Industrial Engineering, vol. 149, p. 106785, November 2020.

[11] C. Muenzberg, K. Michl, H. Heigl, T. Jeck, and U. Lindemann, "Further development of TRIZ function analysis based on applications in projects," in International Design Conference - DESIGN 2014, Dubrovnik, Croatia, May 19-22 2014, pp. 333-342.

[12] L. Quoc and M. Tomas, "Distributed representations of sentences and documents," in The 31st International Conference on Machine Learning, Beijing, China, June 2014, vol. 32, no. 2: PMLR, pp. 1188-1196.

[13] E. Tostrup and S. Mesic, "Massive patent data mining," Master's thesis, Department of Computer Science, LTH, Lund University, Sweden, 2019.

[14] J. Kim and S. Lee, "Patent databases for innovation studies: A comparative analysis of USPTO, EPO, JPO and KIPO," Technological Forecasting and Social Change, vol. 92, pp. 332-345, March 2015.

[15] C. K. Chan, K. W. Ng, M. C. Ang, C. Y. Ng, and A. Kor, "Sustainable product innovation using patent mining and TRIZ," in Advances in Visual Informatics, Lecture Notes in Computer Science, H. Badioze Zaman et al. Eds. Cham: Springer, 2021, pp. 287–298.

[16] M. Ghane, M. C. Ang, R. A. Kadir, and K. W. Ng, "Technology forecasting model based on trends of engineering system evolution (TESE) and big data for 4IR," presented at the 2020 IEEE Student Conference on Research and Development (SCOReD), 27-29 September, 2020.

[17] R. Řehůřek. "Doc2Vec Model — gensim." https://radimrehurek.com/gensim/auto_examples/tutorials/run_doc2vec_lee.html (accessed 3 March, 2021).

[18] D. Féron, Corrosion behaviour and protection of copper and aluminium alloys in seawater. Padstow, Cornwall: CRC Press, 2007.

[19] J. Zhao et al., "Effect of surface passivation on corrosion resistance and antibacterial properties of Cu-bearing 316L stainless steel," Applied Surface Science, vol. 386, pp. 371-380, November 2016.

[20] J. Vimala, M. Natesan, and S. Rajendran, "Corrosion and protection of electronic components in different environmental conditions - an overview," The Open Corrosion Journal, vol. 2, pp. 105-113, October 2009.

[21] A. Tiwari, S. Seman, G. Singh, and R. Jayaganthan, "Nanocrystalline Cermet Coatings for Erosion–Corrosion Protection," Coatings, vol. 9, no. 6, p. 400, June 2019.

[22] D. Burduhos Nergis, P. Vizureanu, and C. Bejinariu, "Phosphate surface treatment for improving the corrosion resistance of the C45 carbon steel used in carabiners manufacturing," Materials, vol. 13, no. 15, p. 3410, August 2020.