INTERNATIONAL CONFERENCE ON ENGINEERING DESIGN, ICED21
16-20 AUGUST 2021, GOTHENBURG, SWEDEN

DATA-DRIVEN CREATIVITY: COMPUTATIONAL PROBLEM-EXPLORING IN ENGINEERING DESIGN

Obieke, Chijioke C.;
Milisavljevic-Syed, Jelena;
Han, Ji

University of Liverpool

ABSTRACT

Creativity is required in engineering design. It is required in the aspects of problem-solving - conceptualizing a new solution to a problem, and problem-exploring - conceptualizing a new problem. Studies show that, in both aspects, creativity is a difficult task in practice. The aim of this study is to support the engineering design community by easing the difficulty in the problem-exploring practice. To achieve this, a computational problem-exploring (CPE) model is developed to mimic how design engineers identify a valid design problem. Consequently, a CPE tool - Pro-Explora V1 is developed based on the CPE model. The CPE model consists of a synergy of emergent computational technologies including data retrieval and machine learning. A Markovian model is employed in the CPE model to enable a data-driven random process for exploring design problems. In pilot test, Pro-Explora V1 generated some engineering design-related problems which are meaningful, unique, and could not be distinguished from naturally generated ones. It provides support to design engineers in problem-exploring at the early stage in engineering design. This study contributes to the global effort towards data-driven processes in the fourth industrial revolution.

Keywords: Industry 4.0, Creativity, Design engineering, Problem-exploring, Markov chain model

Contact:
Han, Ji
University of Liverpool
Industrial Design
United Kingdom
ji.han@liverpool.ac.uk

Cite this article: Obieke, C. C., Milisavljevic-Syed, J., Han, J. (2021) 'Data-Driven Creativity: Computational Problem-Exploring in Engineering Design', in *Proceedings of the International Conference on Engineering Design (ICED21),* Gothenburg, Sweden, 16-20 August 2021. DOI:10.1017/pds.2021.83

1 INTRODUCTION

Boden (2004) defines creativity as the ability to come up with something new and useful. In almost all its occurrences, the word 'creativity' connotes idea generation for problem-solving (PS) in engineering design. However, problem-exploring (PE) as an activity of searching and discovering a new problem in engineering design is within the scope of creativity. In fact, as studies show, PE is an important but disregarded aspect of creativity in engineering design (Obieke et al., 2020; Einstein and Infeld, 1938). To an extent, for decades, this disregard is influenced by the many definitions and descriptions of creativity which focus only on idea generation for PS (Kaplan, 2019; Plucker et al., 2004). Getzels (1979) points out the importance of new problems in creating opportunities for new solutions. Despite the importance of PE in engineering design, its practice remains significantly low compared with PS. Generally, conceptualizing new ideas to solve a problem is a difficult, challenging, and time-consuming task in practice (Grigorenko, 2019; Nicholl and McLellan, 2007). However, studies suggest that conceptualizing a new problem is comparatively a more difficult, challenging, and time-consuming task in practice (Harris and Zeisler, 2002; Fischer, 1994; Einstein and Infeld, 1938). This gives an insight why PE is not widely practiced as PS in engineering design. In addition, there are support tools for PS such as TRIZ, Idea Inspire, Combinator and so on (Han et al., 2019). However, there are no equivalent tools to support PE in engineering design.

The aim in this study is to explore the use of a Markovian model in developing a data-driven computational tool to support PE in engineering design. The tool is used, at the early stage in engineering design, to support both experienced and novice design engineers in PE. The emergent artificial intelligence (AI) technologies in the fourth industrial revolution (4IR) inspire research thinking towards digitalization or digital data-driven activities and processes (Chiarello et al., 2020; Milisavljevic-Syed et al., 2019). This provides the research opportunity in this study to support the PE aspect of creativity with the AI technologies especially big data, data retrieval, natural language processing (NLP), machine learning and duplication recognition (Nitta and Satoh, 2021). In this study, the Markovian model is used in synergy with these AI technologies to computationally support PE in engineering design. A similar hypothetical approach would be a synergy of BERT (bidirectional encoder representations from transformers) technology and Markovian model (Wang and Cho, 2019). However, relatively, this alternative would require data training, a higher computational power, time, speed, and cost. This is due to the long-term persistence of information in BERT network. Prior to now, it is believed that identifying a problem could be impossible using computational means (Celik, 2019). This study contributes to the current global efforts toward digitalization or data-driven processes (Müller and Trahasch, 2019).

In the following section, digitalization and its importance in the PE aspect of creativity are discussed. In Section 3, the methodology adopted in this study to facilitate computational problem-exploring (CPE) in engineering design is presented. This includes the CPE model and tool developed which is based on the Markov chain model (MCM) and hidden Markov chain model (HMCM). The initial test results and verification for the CPE tool - Pro-Explora V1, are discussed in Section 4. In the last section, the conclusion of the study is presented.

2 DIGITALIZING CREATIVITY IN ENGINEERING DESIGN IN THE 4IR

Creativity remains important and is a key focus, both in the industry and academia, as development moves into the 4IR (Hecklau et al., 2016). The dominant technologies from the first to 4IR are mechanical power, electrical power, invention of computers (digitization), and digitalization, respectively (Preuveneers and Ilie-Zudor, 2017). Digitization - the conversion of information or data from one form, usually analog form, to digital form (Gobble, 2018), largely occurred in the third industrial revolution. When data is in a digital form, computers can interpret, store, process, and transmit it. The concept of digitalization - an automated digital data-driven process involving decision-making, becomes prominent in the 4IR. Bloomberg (2018) describes digitalization as the conversion of a conventional process to a digital process. This is possible due to the emergent AI technologies of the 4IR particularly big data, information retrieval, natural language processing, machine learning and duplication recognition. Albeit the imminent applications of these AI technologies are not fully known, engineering design is identified as one area of application (Crawford, 2018). Within engineering design, the conventional creativity practice would be impacted by these mentioned AI technologies. However, a comprehensive understanding of creativity in engineering design is vital as presented in section 2.1.

https://doi.org/10.1017/pds.2021.83 Published online by Cambridge University Press

2.1 Understanding creativity from the problem-exploring aspect in engineering design

As mentioned in the introduction, the concept of creativity is mostly associated to PS. The focus in academia and industry is on educating and training a design engineer on employing creativity for PS (Kaplan, 2019; Tekmen-Araci and Mann, 2019). This conflicts with the expectations for design engineers, especially in the 21st century. A design engineer is expected to have the ability to: 1) provide a solution to a problem, and 2) conceptualize or discover a new problem (Gangopadhyay, 2014; NAE, 2004). These expectations are becoming more pronounced as global development in the 21st century transits to the 4IR (Charette, 2017).

There are implications for the continuous propagation of the understanding of creativity around PS only while disregarding PE. One of such implications is that the society would consist of design engineers who are educated or trained to generate ideas for solving problems and not seeking to identify or explore new problems. Another implication would be a declining number of inventions compared to innovations as discussed in Section 2.2.

2.2 Why does problem-exploring aspect of creativity deserve attention?

The definition of creativity in the 4IR should reflect its importance in PE and PS (Obieke et al., 2020). A design problem could be defined as an occurrence, observation, a description, or conception with a potentially useful engineering design solution which is not available or readily available (Runco, 2014; Krulik and Rudnick, 1987). The identification of a new design problem in engineering design could lead to inventions, advancement of science, and new creative solutions or concepts as illustrated in Figure 1.

		A	B
		Created/Discovered Problem	**Presented Problem**
1	**Unknown Solution**	Invention/Creativity	Innovation/Creativity
2	**Known Solution**	Knowledge/Creativity	Knowledge/Experience

Figure 1: Engineering design problem-exploring and problem-solving matrix

In Figure 1, three types of problems are indicated - presented, discovered, and created problems. According to Getzels (1979), a presented problem refers to a problem that already has an existing formulation, method of solution and solution which may or may not be known to the person whom the problem is given. A discovered problem refers to an existing problem which may or may not have an existing formulation, method of solution, or solution. Unlike a presented problem, a discovered problem is identified by oneself instead of being given (or presented) to one. A created problem refers to a problem that is never considered as a problem, and may not constitute a threat to human existence, until someone conceptualized and translated it to make it an apparent problem. It is important to note here that a solution that does not yet exist could be regarded as a problem. Solving a discovered or created problem with an unknown solution involves invention and creativity in engineering design (*A1* in Figure 1). To conceptualize or discover a new problem requires creativity, albeit such problem could be solved with the knowledge of existing solutions (*A2* in Figure 1). Solving a presented problem involves just experience and knowledge since a presented problem has a known solution (*B2* in Figure 1). However, to solve a presented problem with an unknown solution involves creativity to conceptualize the innovative or unknown solution (*B1* in Figure 1). Studies show that PE is a difficult task (Jørgensen, 2006; Yoshioka et al., 2005). This hints why PE is not widely practiced in engineering design in addition to lack of motivation due to no specific support tool (Obieke et al., 2020).

In this section, the importance of PE aspect of creativity in engineering design is discussed. In the next section, how the emergent AI technologies are used to support PE in engineering design is presented.

3 COMPUTATIONAL PROBLEM-EXPLORING USING MARKOVIAN MODEL

In this section, an approach that integrates a Markovian model and emergent AI technologies to support PE is developed. The natural PE process and Markovian model are investigated. A CPE model is developed to computationally mimic the natural PE process. Subsequently, a CPE tool - Pro-Explora V1, is developed based on the CPE model.

https://doi.org/10.1017/pds.2021.83 Published online by Cambridge University Press

3.1 Models for natural problem-exploring

The natural process for identifying a new design problem is discussed in this section. Alter and Dennis (2002) mention two natural models or processes through which new and inspiring problems can be identified. One is the "rational model" which is a formal model of science in which a new problem is identified and selected through the careful analysis of previous study and theory. Through this analysis a prompt or accidental discovery of a problem is possible (Polanyi, 1958). The other model - "garbage can model," asserts that new problems are identified and selected as opportunities arise through future needs, emergent technologies, and other fields. The garbage can model implies that a new problem can emerge from random connections of existing problems. This suggests that the natural PE process is not orderly as some authors also point out (Dennis and Valacich, 2001; Cohen et al., 1972). A conventional approach used by designers to explore and verify a new design problem is given in Table 1.

Table 1: Natural and computational equivalent approaches for problem-exploring

	Natural approach	Computational approach
1	Accidentally identify or deliberately frame a potential problem or idea through random observations, social interactions, analyses of previous studies, attentional and conceptualization abilities.	Autonomously frame a potential problem or idea through random connections in big data (social media, websites, etc.) on previous studies using data retrieval, NLP, machine learning, Markovian model, and coding capabilities.
2	Perform sequential autonomous search for prior existence in relevant databases by using search engines.	Perform parallel autonomous search for prior existence in relevant databases using similarity and duplication recognitions.
3	Make a decision based on search result.	Make a decision subject to a designer's acceptance.

In Table 1, the basic natural approach used by design engineers in identifying a valid problem in engineering design are outlined. The computational equivalent of this natural approach is given in the last column of the table. Most intellectual property offices (IPO) endorse the natural approach in Table 1 for verifying the originality of a problem or solution (IPO UK, 2017). In Section 3.2, the use of the Markovian model to computationally simulate the natural approach in Table 1 is explored.

3.2 Markovian model for engineering design problem generation

The term 'Markovian' describes a process that exhibits the Markov property. The concept of the Markov property comes from the Markov chain model (MCM) and hidden Markov chain model (HMCM). In this study, this concept is used to develop a data-driven computational PE model. This is done in synergy with some computational technologies which include big data, data retrieval, NLP, machine learning, and duplication recognition. The resulting computational model mimics the natural PE process presented in Table 1. A MCM represents a specific type of random process (Meyn and Tweedie, 2009), which does not retain the memory of its past states (Norris, 1997), as shown in Figure 2. The assumption in the MCM is that, given a present state, the probability of the next state only depends on the present state and not on past state(s) (Sheskin, 2011).

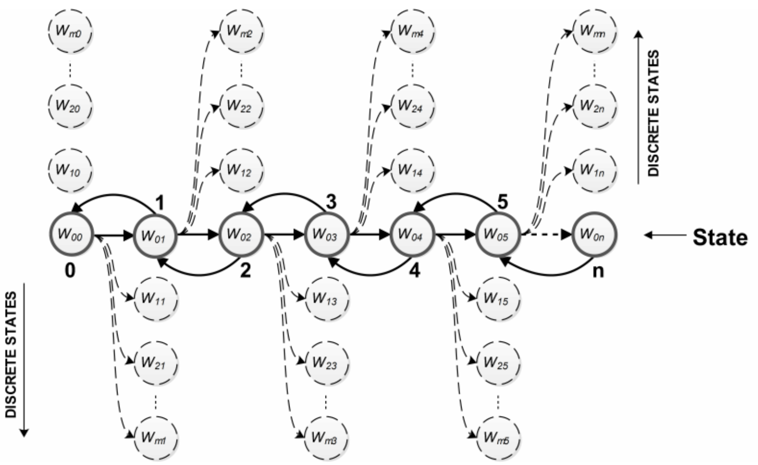


Figure 2: Engineering design problem generation (DPG) model

https://doi.org/10.1017/pds.2021.83 Published online by Cambridge University Press

A model of engineering design problem generation is presented in Figure 2. It contains an indexed sequence of random variable states, W_{00}, W_{01}, W_{02}, W_{03}, ..., W_{0n}, where $n \in \mathbb{Z}^{+}$. For clarity, $\mathbb{Z}^{+} = \{0, 1, 2, 3, \dots\}$. A state in a random sequence represents a likely value for a random variable in the sequence. In Figure 2, the next state in the sequence depends on the present state only. This type of dependency is referred to as the Markov property. It is useful in explaining various real-life situations requiring decisions such as weather forecasting and machine maintenance. The HMCM is an extension of the MCM. It is a two-stage random process with hidden states and physically observable states (Plötz and Fink, 2011). According to Rabiner (1989), the application of MCM to a real-life problem has limitations. It does not fully represent the intent when used in many problems. The HMCM is used to overcome this limitation. Hence, in many real-life cases, the MCM and the HMCM can be used. The MCM is extensively used in handwriting recognition, and learning models of sequential data such as words (Plötz and Fink, 2011). However, the application of MCM in PE in engineering design is not explored. Specific adjustments would be required to apply MCM in PE to generate meaningful concepts with useful engineering design applications (Fink, 2014), as done in this study.

Applying the two-stage HMCM to the DPG model in Figure 2, the sequence W_{00}, W_{01}, W_{02}, W_{03}, ..., W_{0n} represents the physically observable states (first stage) of the two-stage HMCM. The observable states depend on the probability of the hidden or non-physically observable states (second stage) of the two-stage HMCM. These hidden states are the vertical or column lists of finite mutually exclusive discrete states in Figure 2, with the circles bordered with broken outlines. For example, at index 3 of the sequence of states in Figure 2, the hidden states are W_{13}, W_{23}, ..., W_{m3}, where $m \in \mathbb{Z}^{+}$. For clarity, $\mathbb{Z}^{+} = \{0, 1, 2, 3, \dots\}$. An assumption is that the transition of the physically observable states in Figure 2 are observed at equal time intervals (time homogenous) at the indices, *0*, *1*, *2*, *3*, ... *n*, known as epochs (Privault, 2013). Assuming Markov property applies, then each observable state presented in Figure 2 is memoryless on past states. For example, using W_{04} as the present state, the probability of the next state W_{05} only depends on the present state W_{04} and not on past states say W_{03}, W_{02} and so on. If the state at epoch n in Figure 2 is S_i (that is, $W_{0n} = S_i$), then let $P(W_{0n} = S_i)$ represent the probability that at epoch n the Markov chain (sequence) is in state S_i. Similarly, the probability that at epoch $n + 1$ the chain is in state S_j can be represented as $P(W_{0(n+1)} = S_j)$. Hence, starting from a present state, the probability of the next state only depends on the present state and not on past states as expressed in Equation 1.

$$p_{ij} = P(W_{0(n+1)} = S_j \mid W_{0n} = S_i) \tag{1}$$

The term p_{ij} denotes a conditional or transition probability (Rabiner, 1989). As assumed, the transition probabilities are time homogenous or do not change with time such that they do not depend on the epoch n. So, Equation 1 can be written as,

$$p_{ij} = P(W_{0(n+1)} = S_j \mid W_{0n} = S_i) = P(W_{01} = S_j \mid W_{00} = S_i) \tag{2}$$

The matrix that contains the transition probabilities p_{ij}, in Equation 2, is known as a one-step transition probability matrix. This matrix, P, is expressed in Equation 3 and can represent a random process.

$$P = [p_{ij}] = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1n} \\ p_{21} & p_{22} & p_{23} & \cdots & p_{2n} \\ p_{31} & p_{32} & p_{33} & \cdots & p_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{m1} & p_{m2} & p_{m3} & \cdots & p_{mn} \end{bmatrix} \quad 1 \leq i, j \leq N\,;\ N \in \mathbb{Z}^{+};\ \mathbb{Z}^{+} = \{0, 1, 2, 3, \dots\} \tag{3}$$

In Equation 3, each row represents the present state while each column represents the next state or transition. Equation 3 represents an N x N square matrix ($P_{mn} = P_{mm} = P_{nn}$) because all the possible states for a Markov chain with N states will constitute an N x N square matrix. Hence, in Figure 2, the matrix that contains all the possible transition probabilities for the sequence will always be a square matrix irrespective of the number of states in the sequence. In some cases, the initial state probability for a random process may be necessary, for instance when the state of the process is of interest after n transitions. Note that state transitions occur if, and only if, $p_{ij} > 0$. Let $p_j^{(0)} = P(W_{00} = j)$ represent the probability that at epoch *0* the initial state probability in a Markov chain is j. Then, for a Markov chain with n states, all the initial state probabilities can be represented as a row vector of initial state probabilities as expressed in Equation 4.

$$p_j^{(0)} = [p_1^{(0)} \ \ p_2^{(0)} \ \ p_3^{(0)} \dots p_n^{(0)}] = [P(W_{00} = 1) \ P(W_{00} = 2) \ P(W_{00} = 3) \dots P(W_{00} = n)] \tag{4}$$

https://doi.org/10.1017/pds.2021.83 Published online by Cambridge University Press

Referring to Figure 2, for simplicity, let the states at epochs *0, 1, 2, …, n* be represented as *j, k, l, …, z,* respectively. The probability of the states at the respective epochs would be $P(W_{00} = j, W_{01} = k, W_{02} = l, \ldots, W_{0n} = z)$. Based on Markov property, the joint probability of the state transitions or sample path can be determined as:

$$\left.\begin{aligned} &P(W_{00} = j, W_{01} = k, W_{02} = l, \ldots, W_{0n} = z) \\ &= P(W_{00} = j)P(W_{01} = k \mid W_{00} = j)P(W_{02} = l \mid W_{01} = k) \ldots \\ &= p_j^{(0)} p_{jk} p_{kl} \ldots \end{aligned}\right\} \tag{5}$$

If the initial state $p_j^{(0)}$ is known then $P(W_{00} = j) = p_j^{(0)} = 1$. Equation 6 is gotten from Equation 5. Thus,

$$\left.\begin{aligned} &P(W_{01} = k, W_{02} = l, \ldots, W_{0n} = z \mid W_{00} = j) \\ &= P(W_{01} = k \mid W_{00} = j)P(W_{02} = l \mid W_{01} = k) \ldots \\ &= p_{jk} p_{kl} \ldots \end{aligned}\right\} \tag{6}$$

3.3 Modelling the computational problem-exploring process framework

Explained in this section is how the emergent 4IR AI technologies - big data, data retrieval, NLP, machine learning, and duplication recognition, are applied to support PE in engineering design. Studies show that PE can start from the analysis of a previous project or research (Section 3.1). A project title is important and perhaps the most important line of words in any engineering design project (Langford and Pearce, 2019; Oermann and Leonardelli, 2013). It describes the problem addressed in a very much abridged form (Greenspan, 2016). A design project title (problem description) attracts attention and inspires thoughts towards valuable solutions (Hays, 2010). In this study, engineering design project titles are observed to exhibit the Markov property. There is a transition probability for the next word relative to the previous word as presented in Figure 2. This is such that the present word influences the next word only and past words are ignored. Scrapy-based **data retrieval** technology is used in this study to retrieve engineering design-related project titles from the web pages of a website. How the corpus of data is retrieved, cleaned, structured, and used to support PE in engineering design are described using the CPE model presented in Figure 3. The CPE model integrates the DPG model shown in Figure 2.

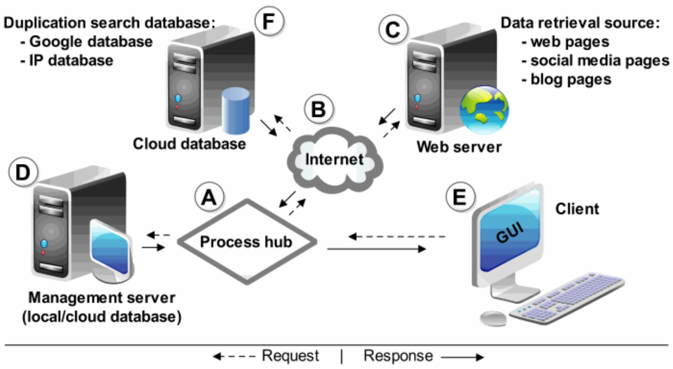


Figure 3: A CPE model for engineering design problems

In Figure 3, a data-driven CPE model for generating engineering design problem is shown. The model simulates the natural PE approach in Table 1. The process hub (*A* in Figure 3) is a computational process in which computational decisions are made using a synergy of emergent computational technologies and python codes. The decisions are towards generating a new design problem based on the DPG model in Figure 2. The process hub initiates an http request, facilitated by scrapy **data retrieval** technology, to retrieve digital data from specified online sources. This follows the path *A-B-C* in Figure 3. The http request from the process hub is sent via the internet (*B* in Figure 3) to the appropriate online sources/servers (*C* in Figure 3). Data from these online sources constitute **big data**. The data used in this study are engineering design-related project titles retrieved from the web pages of a website - finaphd.com. A total of 137 design-related project titles are retrieved. The data and response for the http request are sent back to the process hub via the internet. This follows the path *C-B-A* in Figure 3. During the data retrieval, data cleaning is performed to expunge the non-text, html tag, and white space contents in the data. These contents are not desirable or useful. The data cleaning - a **NLP** technique, is performed using the scrapy data retrieval technology built-in *pipeline* architecture (Hajba, 2018).

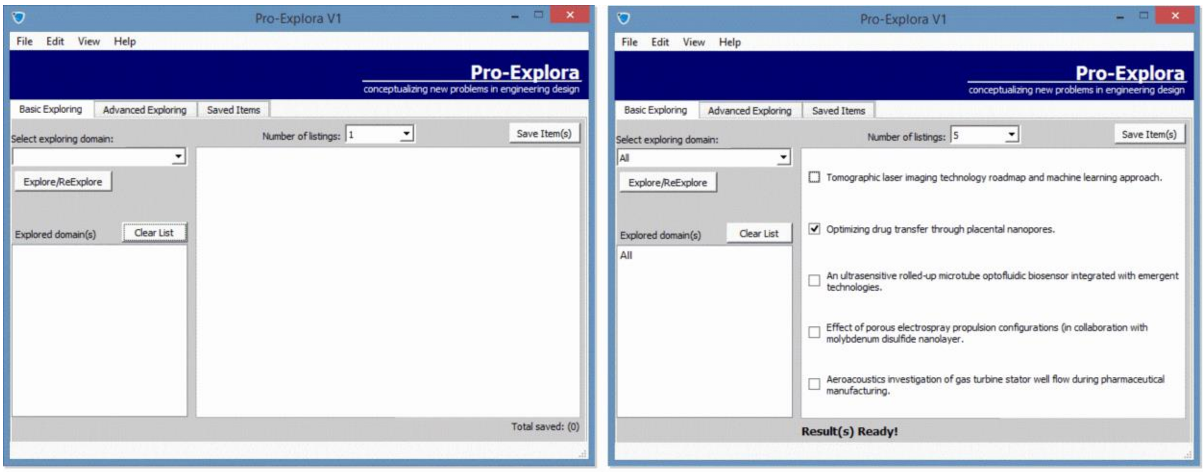
https://doi.org/10.1017/pds.2021.83 Published online by Cambridge University Press

The retrieved data is processed further in the process hub to the required data structure. It is split into dictionary *keys* and *values* data structure using python codes. Specifically, the python code is written to 'learn' (**machine learning**) from the design problems in the corpus retrieved and autonomously generate new ones. Each word in the corpus becomes a *key* and the next word after every occurrence of the *key* becomes a *value* for the *key*. The lists of *values* for the *keys* form the discrete states shown in Figure 2 except the initial states - W_{10}, W_{20}, …, W_{m0}. The first word in each of the retrieved project titles is selected to form the list of the mutually exclusive discrete initial states - W_{10}, W_{20}, …, W_{m0}, in the DPG model in Figure 2 (see Equations 4 - 6). The originally retrieved corpus and the processed python dictionary data structure are saved in the management server (*D* in Figure 3) following the path *A-D*. Note that the process hub, as a computational decision process, runs from within the server (*D* in Figure 3). It is distinctly represented in Figure 3 for the convenience of explanation. Apart from the initial states, all other discrete states shown in Figure 2 are adjusted by adding new words while maintaining existing context and parts of speech. The contextual synonyms of words ending with a full stop in the *values* list are added to the list containing the words. For example, some of the contextual synonyms added to the list containing the word "extractor." are - "conveyor.", "pump.", "squeezer.", "unpacker.", and "juicer." This domain application-specific adjustment is necessary for a useful result when employing a Markovian model in real-life (Rabiner, 1989).

A new design problem is generated using the structured data stored in the management database and based on the DPG model in Figure 2. This is initiated through a simple graphical user interface (GUI) or client (*E* in Figure 3). An initial word is selected, and the following words are selected randomly from the words stored in the *values* list until set conditions are met. The conditions are - selecting a word that ends with a full stop and generating a problem of 6 to 12 words (see Section 4). A **duplication recognition** search is used to ensure that there is no prior existence for a new generated design problem. The duplication search is done in the original retrieved corpus and can be extended to cloud databases (*F* in Figure 3). The cloud database could be a google, intellectual property (IP), and/or similar databases. A generated design problem with a duplicate is discarded and a new one is generated. The process repeats until a design problem with no duplicate is identified for a design engineer's (user's) acceptance or rejection. This way a new and valid engineering design problem is easily identified or inspired. A CPE tool - Pro-Explora V1, developed based on the CPE model is presented in Section 4.

4 DISCUSSIONS

The pilot test and results obtained for Pro-Explora V1 - a CPE model-based tool, are discussed in this section. The default GUI of the tool as seen by a design engineer (*E in Figure 3*) is shown in Figure 4a.



(a) (b)

Figure 4: Pro-Explora V1 GUIs; (a) Default GUI, (b) GUI with a selected problem to save.

In Figure 4a, the preferential options for generating a design problem from Pro-Explora VI default GUI are shown. In Figure 4b, five design problems generated by Pro-Explora V1 are shown and can be selected and saved. They are successfully generated based on the CPE model in Figure 3. For portable screen view, the maximum number of problems to display at once is set to 5. Based on the retrieved data size used during the pilot test, Pro-Explora V1 generated about 100 varied design problems per minute.

https://doi.org/10.1017/pds.2021.83 Published online by Cambridge University Press

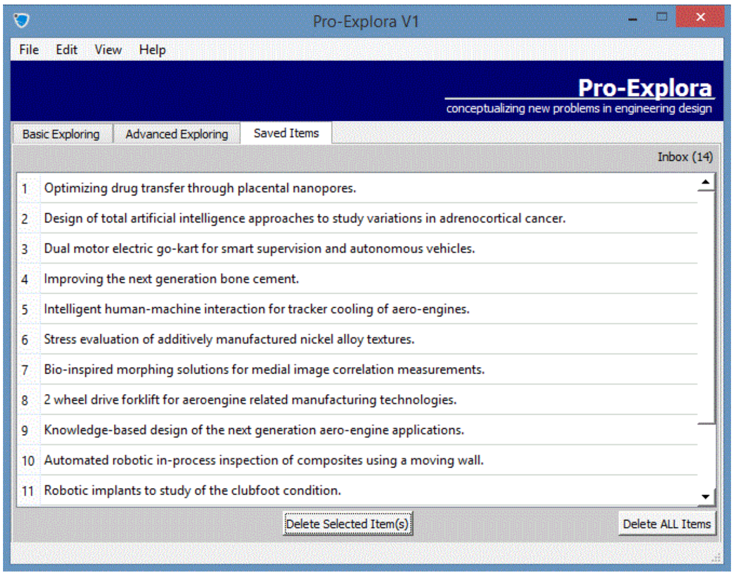


Figure 5: List of saved problems generated by Pro-Explora V1.

A list of some Pro-Explora V1 generated and saved design problems are shown in Figure 5. The number of words (n) for each design problem is constrained to a maximum of 12 and minimum of 6, that is $6 \leq n \leq 12$. This is due to the widely suggested rule of 12 words maximum for an inspiring project title or problem description (Bahadoran et al., 2019; Hays, 2010). The initial word for each design problem is selected from a list of mutually exclusive discrete words described in Section 3.3 and with Equations 4 to 6. The verification of the pilot test results of Pro-Explora V1 is focused on two aspects: 1) the validity and, 2) uniqueness of the generated design problems as discussed in Section 4.1.

4.1 Verification of Pro-Explora V1: Pilot test

As a pilot case study, the verification of the results of Pro-Explora V1 is performed using two professionals with engineering design-related backgrounds as participants. To guard against bias (Jordanous, 2012), five of the saved design problems generated by Pro-Explora V1 (Figure 5) are selected and presented to the participants alongside naturally generated ones. The design problems are selected by inspection on how natural they appear. The selection of five design problems is to keep the case study within 10 minutes so that participants will be willing to take part. The participants were unable to distinguish between the naturally generated problems and Pro-Explora V1 generated problems. This shows a significant success as it confirms the validity of Pro-Explora V1 results. According to Colton and Wiggins (2012), "in many application domains, it is a significant milestone when observers cannot reliably distinguish between a computer generated artefact and one produced by a person." A second verification of Pro-Explora V1 results is based on uniqueness. As verified, the duplicates of the generated design problems are not in the original corpus used in generating the problems. This shows uniqueness because duplication recognition search is incorporated in the CPE model in Figure 2.

The results obtained in this pilot test is enabled by the approach adopted in this study and adjustment made in employing the two-stage HMCM (Section 3.3). Hence, this study is a step in the right direction towards digitalizing PE in engineering design. The nature of the retrieved corpus used for the pilot test of Pro-Explora V1 is of significance to the results obtained as discussed in Section 4.2.

4.2 Description of Pro-Explora V1 results

The corpus used to generate the design problems for the pilot test of Pro-Explora V1 (Figure 5) is retrieved from various engineering design-related fields. This accounts for the nature of results obtained which cuts across medical, artificial intelligence, mechanical and electrical engineering fields, and so on. Thus, these results reflect the various fields from which the corpus is retrieved. These fields are not specifically sorted from the corpus retrieval source. Hence, the sorting of the retrieved corpus into specific engineering design domains is not done in this study. Generating a problem from these design interrelated fields could expose or inspire a rare design problem based on the 'garbage can' model (Section 3.1). However, when the intent is to explore for a design problem from a specific engineering design domain then retrieving a large corpus from that specific domain would be advantageous.

https://doi.org/10.1017/pds.2021.83 Published online by Cambridge University Press

5 CONCLUSIONS

In this study, a CPE model is developed to assist in identifying valid, useful, and unique engineering design problems. Consequently, a CPE tool - Pro-Explora V1 is developed based on the CPE model. The purpose of this tool is to support design engineers in the difficult creativity task of PE. Studies show that PE, as an important part of engineering design, is not widely practiced in academia and industry. PE is necessary for inventions, and advancement of science. Some contributing factors to the scarce practice of PE include the lack of a support tool and difficulty involved in PE. The CPE model developed in this study mimics how design engineers identify a valid design problem in real-life. It employs: 1) a Markovian model comprising a two-stage adjusted hidden Markov model which enables a data-driven random process for PE, and 2) a combination of emergent computational technologies which includes big data, data retrieval, NLP, machine learning and duplication recognition. The availability of big data enables a corpus of naturally generated engineering design problems to be retrieved and used as 'learning' data for the data-driven random process. On pilot test, Pro-Explora V1 generated some engineering design-related problems which are meaningful, unique, and could not be distinguished from naturally generated ones when presented to participants. Also, none of the Pro-Explora V1 results exists in the original corpus used as training data to generate the new design problems. This is due to the duplication recognition search concept incorporated in the CPE model.
Pro-Explora V1 provides support to design engineers in PE at early stage in engineering design. This study contributes to the global effort towards data-driven processes in the 4IR. A further development of Pro-Explora V1 is to conduct a second case study with larger participants and extend the duplication recognition search to google database. This would provide more user experience and opinions. The information will be used in optimizing Pro-Explora V1, and the relevance and usefulness of the output specifically evaluated. The long short-term memory (LSTM) aspect of deep learning technology would be explored in developing Pro-Explora V1 further. The success of Pro-Explora V1 so far is convincing with promising indication of its direct applications in academia and industry.

ACKNOWLEDGMENTS

The Tertiary Education Trust Fund (TETFund) Nigeria is hereby acknowledged for the award of scholarship for this study. The support of the University of Port Harcourt Nigeria is acknowledged.

REFERENCES

Alter, S. and Dennis, A.R. (2002), "Selecting research topics: Personal experiences and speculations for the future", *Communications of the Association for Information Systems*, Vol. 8 No.1 pp. 21.

Bahadoran, Z., Mirmiran, P., Kashfi, K. and Ghasemi, A. (2019), "The principles of biomedical scientific writing", *International Journal of Endocrinology and Metabolism*, Vol. 17 No.4.

Bloomberg, J. (2018), *Digitization, digitalization, and digital transformation: confuse them at your peril. [Online]* Forbes. https://www.forbes.com (17-09-2020).

Boden, M.A. (2004), *The creative mind: Myths and mechanisms*, Routledge, New York.

Celik, Y. (2019), *Rules for Doing and Managing Research Project. [Online]* Google Scholar. https://dergi.biruni.edu.tr/wp-content/uploads/2019/12/2-Yusuf-Celik.pdf

Charette, M. (2017), "Woods Hole Sea Grant in the 21st Century: Issues, Opportunities, Action for Massachusetts 2018-2021",

Chiarello, F., Bonaccorsi, A. and Fantoni, G. (2020), "Technical sentiment analysis. Measuring advantages and drawbacks of new products using social media", *Comput.Ind.*, Vol. 123 pp. 103299.

Cohen, M.D., March, J.G. and Olsen, J.P. (1972), "A garbage can model of organizational choice", *Adm.Sci.Q*.

Colton, S. and Wiggins, G.A. (2012), "Computational creativity: The final frontier?", *Ecai*, Montpelier, pp.21-6.

Crawford, M. (2018), *How Industry 4.0 Impacts Engineering Design. [Online]* The American Society of Mechanical Engineers. https://www.asme.org (05-10-2020).

Dennis, A.R. and Valacich, J.S. (2001), "Conducting experimental research in information systems", *Communications of the association for information systems*, Vol. 7 No.1 pp. 5.

Einstein, A. and Infeld, L. (1938), *The evolution of physics*, Simon & Schuster, New York.

Fink, G.A. (2014), *Markov models for pattern recognition: from theory to applications*, Springer, London.

Fischer, G. (1994), "Turning breakdowns into opportunities for creativity", *Knowledge-Based Syst.*, Vol. 7 No.4 pp. 221-32. https://doi.org/10.1016/0950-7051(94)90033-7.

Gangopadhyay, D. (2014), "The dawn of the creative age: Fostering creativity among engineering students", *Teaching Innovation Projects*, Vol. 4 No.1.

Getzels, J.W. (1979), “Problem finding: A theoretical note”, *Cognitive Science - A Multidisciplinary Journal*, Vol. 3 No.2 pp. 167-172. https://doi.org/10.1207/s15516709cog0302_4.
Gobble, M.M. (2018), “Digitalization, digitization, and innovation”, *Research-Technology Management*, Vol. 61 No.4 pp. 56-9. https://doi.org/10.1080/08956308.2018.1471280.
Greenspan, Y.F. (2016), *A guide to teaching elementary science: Ten easy steps*, Brill Sense, Netherlands.
Grigorenko, E.L. (2019), “Creativity: a challenge for contemporary education”, *Comparative Education*, Vol. 55 No.1 pp. 116-32.
Hajba, G. (2018), *Website Scraping with Python: Using BeautifulSoup and Scrapy*, Springer Science, New York.
Han, J., Park, D., Shi, F., Chen, L., Hua, M. and Childs, P.R. (2019), “Three driven approaches to combinational creativity: Problem-, similarity-and inspiration-driven”, *Proc.Inst.Mech.Eng.Part C*, Vol. 233 No.2 pp. 373
Harris, S.D. and Zeisler, S. (2002), “Weak signals: Detecting the next big thing”, *The Futurist*, Vol. 36 No.6.
Hays, J.C. (2010), “Eight recommendations for writing titles of scientific manuscripts”, *Public Health Nursing*, Vol. 27 No.2 pp. 101-3.
Hecklau, F., Galeitzke, M., Flachs, S. and Kohl, H. (2016), “Holistic approach for human resource management in Industry 4.0”, *Procedia Cirp*, Vol. 54 No. pp. 1-6.
Jordanous, A. (2012), “A standardised procedure for evaluating creative systems: Computational creativity evaluation based on what it is to be creative”, *Cognitive Computation*, Vol. 4 No.3 pp. 246-79.
Jørgensen, U. (2006), “Engineering design competences–controversial relations between techno-science discipline and engineering practice domains pointing to new foundations for engineering knowledge”, *INES Workshop, Virginia Tech*, Citeseer,
Kaplan, D.E. (2019), “Creativity in education: Teaching for creativity development”, *Psychology*, Vol. 10 No.2 pp. 140-7. https://doi.org/10.4236/psych.2019.102012.
Krulik, S. and Rudnick, J.A. (1987), *Problem solving: A handbook for teachers*, Allyn and Bacon, Boston.
Langford, C.A. and Pearce, P.F. (2019), “Increasing visibility for your work: The importance of a well-written title”, *American Association of Nurse Practitioners*, Vol. 31 No.4 pp. 217-8.
Meyn, S. and Tweedie, R.L. (2009), *Markov Chains and Stochastic Stability*, Cambridge University Press, Cambridge. 10.1017/CBO9780511626630.
Milisavljevic-Syed, J., Allen, J.K., Commuri, S. and Mistree, F. (2019), “Design of networked manufacturing systems for Industry 4.0”, Elsevier, pp.1016-21. https://doi.org/10.1016/j.procir.2019.03.244.
Müller, D. and Trahasch, S. (2019), “Architecture of a Big Data Platform for a Semiconductor Company”, *DATA ANALYTICS* 2019, pp. 41.
NAE - National Academy of Engineering, US (2004), *The engineer of 2020: Visions of engineering in the new century*, National Academies Press, Washington DC.
Nicholl, B. and McLellan, R. (2007), “The Contribution of Product Analysis to Fixation in Students’ Design and Technology Work”, *The Design and Technology Association International Research Conference* 2007.
Nitta, K. and Satoh, K. (2021), “AI Applications to the Law Domain in Japan”, *Asian Journal of Law and Society*, pp. 1-24.
Norris, J. (1997), *Markov Chains*, Cambridge University Press, New York.
Obieke, C., Milisavljevic-Syed, J. and Han, J. (2020), “Supporting Design Problem-exploring with Emergent Technologies”, *Procedia CIRP*, Vol. 91 pp. 373-81. https://doi.org/10.1016/j.procir.2020.02.189.
Oermann, M. H. and Leonardelli, A. K. (2013) ‘Make the Title Count’ *Nurse Author &Editor Newsletter*, Vol. 23
Plötz, T. and Fink, G.A. (2011), *Markov Models for Handwriting Recognition*, Springer, London.
Plucker, J.A., Beghetto, R.A. and Dow, G.T. (2004), “Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research”, *Educational psychologist*, Vol. 39 No.2 pp. 83-96.
Polanyi, M. (1958), *Personal Knowledge: Towards a post critical philosophy*, Routledge, London.
Preuveneers, D. and Ilie-Zudor, E. (2017), “The intelligent industry of the future: A survey on emerging trends, research challenges and opportunities in Industry 4.0”, *Journal of Ambient Intelligence and Smart Environments*, Vol. 9 No.3 pp. 287-98. https://doi.org/10.3233/ais-170432.
Privault, N. (2013), *Understanding Markov Chains: Examples and Applications*, Springer-Verlag, Singapore.
Rabiner, L.R. (1989), “A tutorial on hidden Markov models and selected applications in speech recognition”, *Proc IEEE*, Vol. 77 No.2 pp. 257-86.
Runco, M.A. (2014), *Creativity. Theories and Themes: Research, Development, and Practice*, Elsevier, London.
Sheskin, T.J. (2011), *Markov chains and decision processes for engineers and managers*, CRC Press, London.
Tekmen-Araci, Y. and Mann, L. (2019), “Instructor approaches to creativity in engineering design education”, *Proc.Inst.Mech.Eng.Part C*, Vol. 233 No.2 pp. 395-402.
UKIPO - Intellectual Property Office UK (2017), *IP Basics*, Crown Copyright, Newport.
Wang, A. and Cho, K. (2019), “Bert has a mouth, and it must speak: Bert as a markov random field language model”, *arXiv preprint arXiv:1902.04094*.
Yoshioka, T., Suganuma, T., Tang, A.C., Matsushita, S., Manno, S. and Kozu, T. (2005), “Facilitation of problem finding among first year medical school students undergoing problem-based learning”, *Teach.Learn.Med.*, Vol. 17 No.2 pp. 136-41.