Design-by-Analogy: Exploring for Analogical Inspiration With Behavior, Material, and Component-Based Structural Representation of Patent Databases

Hyeonik Song
School of Mechanical Engineering,
Georgia Institute of Technology,
801 Ferst Drive. MRDC 3340,
Atlanta, GA 30332
e-mail: hyeoniksong@gatech.edu

Katherine Fu[1]
School of Mechanical Engineering,
Georgia Institute of Technology,
801 Ferst Dr. MRDC 4508,
Atlanta, GA 30332
e-mail: katherine.fu@me.gatech.edu

Design-by-analogy (DbA) is an important method for innovation that has gained much attention due to its history of leading to successful and novel design solutions. The method uses a repository of existing design solutions where designers can recognize and retrieve analogical inspirations. Yet, exploring for analogical inspiration has been a laborious task for designers. This work presents a computational methodology that is driven by a topic modeling technique called non-negative matrix factorization (NMF). NMF is widely used in the text mining field for its ability to discover topics within documents based on their semantic content. In the proposed methodology, NMF is performed iteratively to build hierarchical repositories of design solutions, with which designers can explore clusters of analogical stimuli. This methodology has been applied to a repository of mechanical design-related patents, processed to contain only component-, behavior-, or material-based content to test if unique and valuable attribute-based analogical inspiration can be discovered from the different representations of patent data. The hierarchical repositories have been visualized, and a case study has been conducted to test the effectiveness of the analogical retrieval process of the proposed methodology. Overall, this paper demonstrates that the exploration-based computational methodology may provide designers an enhanced control over design repositories to retrieve analogical inspiration for DbA practice. [DOI: 10.1115/1.4043364]

1 Introduction

Designers often seek inspiration and direction during ideation and the early stages of the design process. Among various efforts to find such inspiration is design-by-analogy (DbA) [1]. Design-by-analogy involves the retrieval of analogies from a design repository, a "database" of existing design solutions (sometimes simply memory), and the transfer of knowledge from a "source" domain to a "target" domain. To facilitate design-by-analogy in practice, several researchers have studied and developed computational supports to retrieve analogies from electronic patent databases [2,3]. The patent database is deemed an ideal design repository for its innovative ideas across various fields of application and sheer size that grows exponentially worldwide [4]. Unfortunately, our understanding of design-by-analogy practice is inadequate compared to the ever-increasing size of the patent database, restricting designers from being able to utilize the database at its full capacity. To address this research gap, the work presented here uses a computational methodology to explore patents for analogical inspiration, with the goal of facilitating the design-by-analogy practice. Specifically, the work presents following main contributions:

(1) The authors generate and visualize hierarchical repositories of large-scale mechanical design-related patents in which designers can interactively explore for analogical inspiration.
(2) The authors generate component-, behavior-, or material-based hierarchical repositories to provide designers different lenses to influence the way they search for analogies in patent data.
(3) The authors test computational methodology for its ability to assist with the identification of analogical inspiration within the generated design repositories using a case study.

1.1 Prior Studies in Design-by-Analogy. Design-by-analogy has been an active research area [2,5–8], and studies have focused on understanding the effects of analogies on ideation and design outcomes. Many other researchers have contributed to this area, as described next. Linsey et al. explored how various types of representation of information affect the designers' ability to identify, retrieve, and map analogies to design solutions [9]. Tseng et al. studied how analogous information of different levels of applicability to the design problem affects ideation when the design problem has an open-goal [7]. Several studies also focused on methodologies to retrieve analogies including but not limited to: the WordTree method, which retrieves functional analogies by systematically integrating the knowledge of designers and design database [10]; a computational technique to recognize biological analogies using causally related functions derived from semantic information [11]; D-APPS, which provides functional analogies based on a design's product requirements [12]; a number of different patent mining tools [3,13–15]. Most DbA tools, including the prior work reviewed here, use a query-based approach or straightforward input–output method to retrieve analogies. In the case of the WordTree method, designers input a word characterizing a design's functionality and are returned a set of functionally analogical words from various domains, compiled by the algorithm, as source of inspiration [10]. The work presented in this paper is differentiated from the prior work in that our methodology gives designers an entire structured design space for exploration where

[1]Corresponding author.

Contributed by the Computers and Information Division of ASME for publication in the Journal of Computing and Information Science in Engineering. Manuscript received August 9, 2018; final manuscript received March 13, 2019; published online April 26, 2019. Assoc. Editor: Matthew I. Campbell.

they can freely interact with the design solution and discover analogies from various potential sources. The emphasis on the exploration for analogical retrieval is discussed in the next section.

1.2 Exploration for Expert Thinking in Design. Based on the findings of prior research, expert designers have exhibited several cognitive characteristics and abilities during the design process. First, they have been shown to have better spatial memory than novices, being able to process more information [16,17]. Bjorklund found that experts gather more information than novices to ideate design solutions for a given problem [18]. Other researchers found that experts can also cognitively organize the information and conceptualize abstract ideas by viewing the problems objectively [19,20]. Experts have been found to have a systematic approach to design [19], being able to spontaneously adapt to the design constraints [21] and develop heuristics or "rules of thumb" to approach the problem [17,22]. The dynamic, flexible, and systematic characteristics of the expert designers that have been discovered in prior research indicate that the explorative approach is better matched with their cognitive mechanisms. The explorative approach allows the designers to retrieve analogical inspiration by interactively exploring a design repository and autonomously recognizing analogical connections among potential design sources, which could have been ruled out by the query-based algorithm. In addition, the user-controlled explorative approach allows designers to personalize their search for analogies using various analogical properties, creating unique representations of the design repository and/or design problem that could lead to diverse creative design output [23]. Some potential analogical properties that could be employed are introduced in Sec. 2.1.

1.3 Patent Database as Design Repository. The U.S. Patent database has several features that make it a suitable design repository for the design-by-analogy practice. The prior design solutions in the database are valuable knowledge that are deemed "patentable." Patentable ideas can be further defined as ideas that are "useful," meaning that the ideas are functional and operable, and novel. The term "novel" here implies that the ideas have not previously existed before [24]. The database size, already enormous at approximately 10 million patents in 2015 [25], grows continuously in various technical fields and promises designers substantial opportunities to explore for design inspiration in multiple domains. The patent database uses classification systems, such as Cooperative Patent Classification (CPC), to categorize the patents into specific domains for efficient patent-retrieval processes [24]. The characteristics of the patent database not only make it an ideal source of innovation but an efficient means for retrieving analogies.

The vast size of the patent database offers a great opportunity for discovering analogies for design-by-analogy practice, but it simultaneously presents a challenge for effective mining of patents. To address this challenge, many computational tools and methodologies have been studied. Song and Luo integrated the mining of patent texts, citation relationships, and inventor information to retrieve patents for assisting data-driven design [26]. Murphy used a Vector Space Model algorithm to evaluate functional analogies within patents [3]. These works implement different computational approaches to retrieve analogies from the electronic patent database. They all exemplify the importance of computational support to access the patented knowledge in the design repository.

2 Methodology

A structural form of data is essential for providing valuable insights into the data. For instance, Linnaeus's tree structure for biological species and Mendeleev's periodic table for chemical elements led to major scientific advancements in understanding nature [27]. Finding a structural form requires a clustering or categorization of data. In text mining and data mining fields, a popular computational technique used for data clustering is non-negative matrix factorization (NMF) [28,29]. NMF is a topic modeling technique that discovers semantically meaningful topics within a large corpus of documents to aid text mining [30,31]. It has been an active research area in text mining for its practical advantages over other semantic techniques such as latent Dirichlet allocation [32]. One advantage is that NMF generates consistent topic clustering results, assuring that users are returned similar results for multiple runs. Also, numerous matrix computation and optimization studies for efficient NMF computation suggest its competency for the large corpus topic modeling [33–36].

Similar to most semantic techniques, NMF starts with transforming a corpus into a word-document matrix, in which the matrix elements represent the frequency of words (rows) occurring in the patent documents (columns). Mathematically, the word-document matrix is represented by $\mathbf{X}\in\mathbb{R}_{+}^{m\times n}$, where m represents the number of words and n represents the number of documents in a corpus. Given $k \ll \min(m, n)$ as a user-specified number of topics, NMF factorizes the input matrix, $\mathbf{X}$, into two non-negative matrices, namely $\mathbf{W}\in\mathbb{R}_{+}^{m\times k}$ and $\mathbf{H}\in\mathbb{R}_{+}^{k\times n}$ such that $\mathbf{X}\cong\mathbf{WH}$. Here, $\mathbf{W}$ is a word-topic matrix whose ith topic column is represented by the weighted distribution of words. Similarly, $\mathbf{H}$ is a topic-document matrix whose jth document column is represented by the weighted distribution of the respective topics. The matrix decomposition is illustrated in Fig. 1 [29].

In this work, the patent data were processed to generate the cluster structures shown in Appendix A (Figs. 12–14). First, word-document matrices of patents were prepared. Second, cluster structures of the patents were generated using NMF. Lim et al. conducted NMF study with the U.S. patents to study the computational performance of the topic modeling tool [37]. However, the work presented in this study is the first DbA study to use NMF for retrieving analogies for a design practice. Last, visual representations of the structures were generated. All computations for

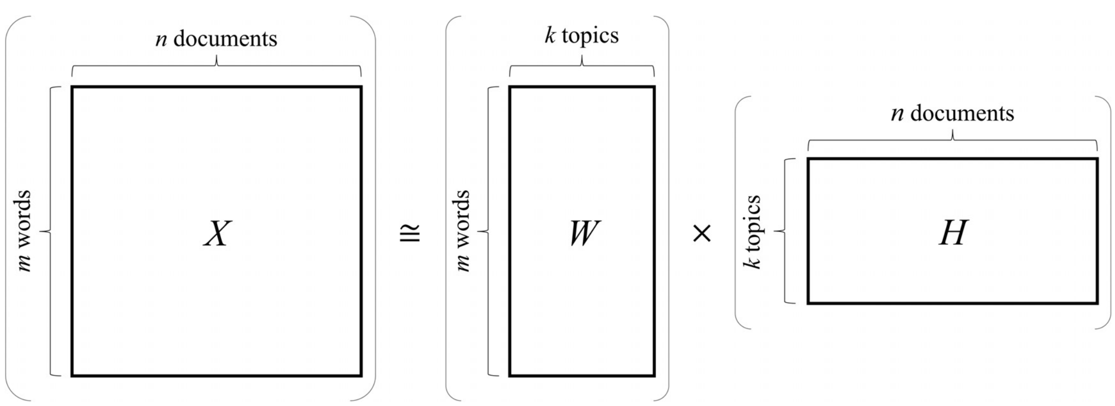

Fig. 1 Illustration of NMF

establishing the algorithmic approach were performed using MATLAB R2016b. Each of these steps is described in detail in the subsequent sections.

2.1 Preparing Word-by-Document Matrices. The first part of the computational approach involves retrieving patent data from a data storage system of United States Patent and Trademark Office (USPTO). The database consists of a bulk of U.S. registered patents, each assigned to at least one classification term called CPC. All registered U.S. patents are categorized into one of nine CPC sections and further assigned into a subsection, which provides a general overview of the patent's design features and area(s) of application. To limit the scope of the study, only mechanical design-related patents were used; Fifty-three CPC subsections were chosen by the researchers as shown in Appendix B. For each subsection, 20 patents were selected using a random number generator, comprising a total of 1060 patents. This sample size (>1000 patents) was chosen to capture various analogical structures in the patent space with a goal of addressing the research gap in implementing the computational DbA tool in a larger-scale design repository. Prior work by the authors has been with sample sizes of 100 patents [38]. In future work, the authors hope to continue to scale up the sample size by orders of magnitude. For each patent document, the researchers used only the words in the abstract, claims, and description sections, as they are the most representative of the patent's design features. All other words in prior patent data and reference sections were omitted as they do not characterize the patent's design aspects.

In addition to the patent documents, a design problem statement was also added to the corpus for generating the word-document matrix. Its purpose was to provide a "starting point" in the cluster space to facilitate data analysis and exploration. The design problem statement, which was used in the researcher's prior study, was as following [38]:

> *Design a device to collect energy from human motion for use in developing and impoverished rural communities in places like India and many African countries. Our goal is to build a low-cost, easy to manufacture device targeted at individuals and small households to provide energy to be stored in a rechargeable battery with approximately 80% efficiency. The energy is intended to be used by small, low power draw electrical devices, such as a radio or lighting device, hopefully leading to an increase in the quality of life of the communities by increasing productivity, connection to the outside world, etc. The target energy production is 1 kWh per day, roughly enough to power eight 25 W compact fluorescent light bulbs for 5 h each per day, or enough to power a CB radio for the entire day. For reference, an average adult human can output about 200 W with full body physical activity for short periods of time, with a significant reduction for sustained power output.*

After the word-document matrix was generated, the matrix was further processed to characterize three design properties—components, behaviors, and materials. In the early stages of the design processes, designers often have diverse objectives and lenses through which they look when searching for inspiration or external information. By allowing for these different lenses to influence the way the design space is structured and inter-related, we give designer the ability to explore in a more tailored and efficient manner than before. The manipulation of the patent data set was done by first manually compiling a list of words that characterize each analogical property, then reducing the original matrix to contain only the rows of the listed words. The word lists were generated by drawing a master wordlist from the patent corpora and classifying each word into one of the three properties. As a designer searches for analogical inspiration, he/she might, for example, ask the following questions when considering components, material, or behavioral content within their potential analogical sources.

- *Component:* What specific components have been integrated to the system/artifact/technology?
- *Behavior:* What are the attributes of the system/artifact/technology that describe how it behaves?
- *Material:* What materials does the system/artifact/technology use or consume?

After a test run, words that appear too frequently were removed to distinguish one patent from another. As a result, wordlists of components (709 words), behaviors (262 words), and materials (377 words) were compiled as shown in Appendices C, D and E. Cohen's kappa inter-rater agreement analysis was performed to verify the robustness of all the wordlists ($\kappa = 0.756$, $p < 0.005$). The refinement does not alter the matrix's numerical element—the frequency of words occurring in each document—but rather removes any words that are "noise" within the given context. This allows designers to explore the patent space using a particular priority, angle, or attribute. A similar practice was done in the author's previous study, in which function and surface features of patents were explored using verbs-only and nouns-only data, respectively [38]. This study is an extension of the prior study in that the components, behaviors, and materials of the patent data are explored to investigate their potential for facilitating design inspiration. For the final step, inverse entropy weighting was performed on the word-document matrices to assign higher weight to words that appear less frequently and vice versa.

2.2 Processing With Topic Clustering Algorithm. The cluster structures of the three patent data sets were generated using NMF. The computational algorithm requires a user-specified number of topics, k, to process the word-document matrix. It is critical that an appropriate topic number is selected for the algorithm, as overly or inadequately clustered data leads to an inaccurate clustering result. Unfortunately, computing an appropriate topic number is still an ongoing research [39], and thus questions the effectiveness of the topic clustering, especially for a large-scale data whose range of topics may vary exceedingly. To cope with this challenge, a computational method similar to Du et al.'s divide-and-conquer non-negative matrix factorization (DC-NMF) was used [33]. As illustrated by a hierarchy structure in Fig. 2, NMF with $k = 2$, or rank-2 NMF, was performed recursively on an input word-document matrix. The rank-2 NMF, which has fast computational speed [40], divides the input matrix into two output matrices of clustered patents. These clusters are then used as inputs for the next iteration. The iteration continues until the processed output cluster contains less than or equal to ten patents. The stopping criterion is an important factor for determining the clustering quality, as it could result in overly or inadequately clustered structures. We acknowledge that the current stopping criterion was experimentally determined and that it generates cluster structures whose qualities are sufficient for analyzing the analogical relationships among patents. However, the stopping criterion of the iterative method is an important area of improvement for an effective topic clustering in future studies.

Throughout the process of generating structures, a label was generated for each cluster so that the analogical relationships among the clustered patents become more transparent and interpretable [38]. In each iteration of NMF, the algorithm outputs **W**, word-topic matrix, and **H**, topic-document matrix. Here, **W** represents the probabilistic distribution of words for each column of topic, implying that the word that has the highest probability score in the jth column in the matrix contributes the most to the jth topic's description. In this study, the columns of words in **W** were sorted in descending order, and the top five words in the column were used to create the cluster label.

2.3 Visualization. The cluster structures of the three patent data sets were visualized using MATLAB's graphing tool to enable the exploration and interpretation of the large-scale cluster space. This section details two visualization methods used to analyze the clustered patent structures.

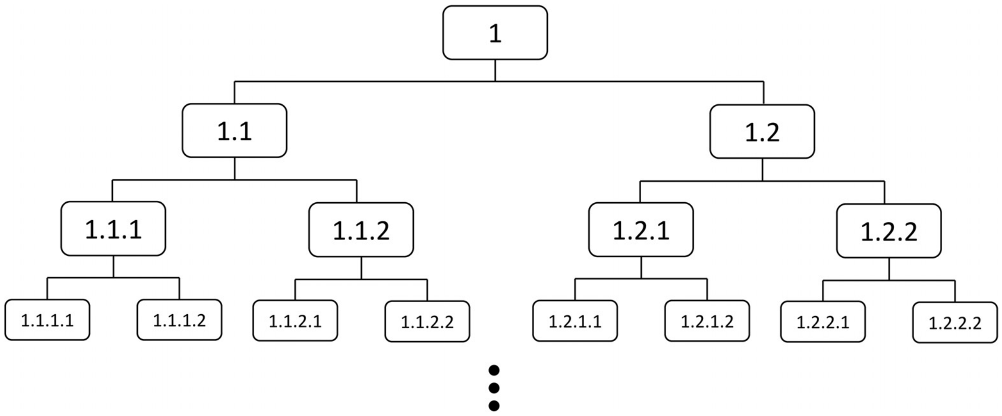

Fig. 2 Illustration of rank-2 NMF iterations

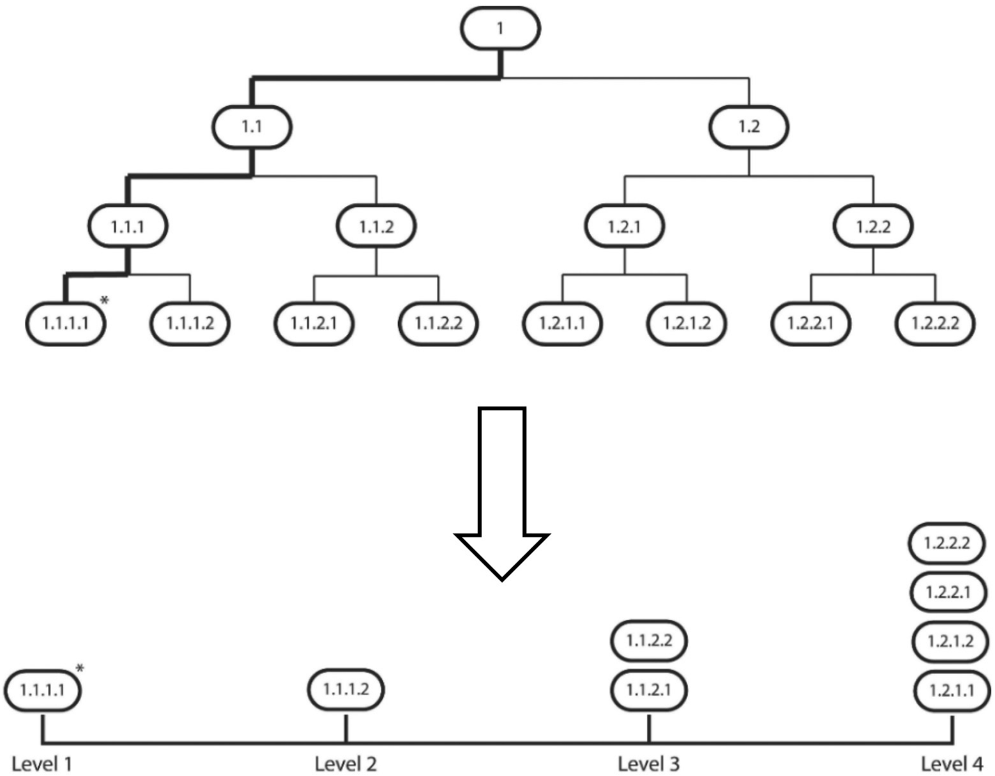

Fig. 3 Transformation of 3D visualization to 2D visualization

2.3.1 Three-Dimensional Hierarchical Visualization. Three-dimensional (3D) hierarchical visualization was used to interactively explore the cluster space. The hierarchy structure is composed of nodes, or clusters of the patent documents, and lines, or the connections between the clusters. The structure contains the initial input data, or "root node" at the center. Starting from the root node, two child nodes branch out recursively outward until all clusters at the end, or "leaf nodes," contain less than or equal to 10 patents. For exploration of the space, the user can rotate or zoom into the structure to search for a node and select the node to view its cluster label and cluster ID number, used for retrieving patent titles. The 3D structure, as shown in Appendix A, was generated using MATLAB's "digraph" function with "force3" layout. The force3 layout generates 3D force directed plot, where the coordinates of nodes and length of edges are determined based on the structure and size of the input graph.

2.3.2 Two-Dimensional Bar Graph Visualization. Figure 3 illustrates the transformation of the three-dimensional hierarchy visualization into a two-dimensional (2D) bar graph visualization. Note that the 3D hierarchy structure in the figure is represented on a 2D graphing space for an effective understanding of the transformation. In the 3D structure, the node of patents iteratively breaks down into two child nodes based on their topic similarity. This implies that the most similar patents in the entire data set would be clustered in a leaf node after a series of NMF iterations. In this manner, if that leaf node is "1.1.1.1" in Fig. 2, the next similar set of patents would be clustered in a leaf node, "1.1.1.2," derived from the same parent node, "1.1.1." Accordingly, the least similar set of patents would be separated in the first iteration performed on the initial node, "1." Once the set of patents is separated, it would go through another series of NMF iterations resulting in several leaf nodes on the other side of the hierarchy structure. This iterative concept was visualized in a two-dimensional bar graph diagram. In the diagram, the horizontal axis represents the level of the hierarchy, equivalent to the series of nodes on the bolded path in Fig. 3. On each level of the hierarchy is a set of leaf nodes with cluster labels that are generated with the separated set of patents at each level. This way, the leaf nodes are distributed across the level of the hierarchy in the order of similarity from a starting leaf node (indicated with an * symbol) in Fig. 3. For consistency of the study, the design problem node—or a leaf node that contains the design problem statement—was chosen as the starting point.

The screenshots of the cluster structures are shown in Appendix A. In the 3D visualizations, the design problem node was selected to display its cluster label and highlight its path from the root node in the center. In the 2D visualizations, all leaf nodes are plotted in the order of similarity to the design problem node on the first level. The two visualization methods were used interchangeably to evaluate the patent data's cluster space. For instance, the 3D visualization was used to view the entire cluster space where individual clusters can be explored by their labels. Two-dimensional visualization was used to sort all clusters on the two-dimensional plane to view the leaf clusters by their similarity to the starting leaf cluster. The analysis results are presented in the next section.

3 Results

The clusters in component, behavior, and material analyses exhibit different characteristics. For the component result, the clusters consist of patents of similar functionality. The functions of the individual components correspond to the subfunctions of the integrated design. The patents in the behavior result are clustered by the design's descriptive quality. The patents in the material result are clustered for two different aspects—(1) material that the design is composed of and (2) material that the design uses or consumes.

The three cluster results are visually unique, suggesting that different design insights can arise from a single patent data set [23]. To confirm this, the researchers first selected three random patents ("pocket tool," "ice gripping sandal," and "shower bath apparatus and spray nozzle"); then, using the computer-generated cluster labels, the researchers evaluated their analogical relationships with other clustered patents as shown in Fig. 4. The figure is a simplified version of 2D bar-graph visualization, where each cluster containing the node number, cluster label, and list of patents is laid out on the order of similarity. In this section, the analogies of the clustered patents are evaluated to understand the clustering mechanism for component-, behavior-, and material-based patent data. However, the meaning of the level of similarity is not discussed due to the limited information to interpret the analogical distance among different clusters. Future studies need to investigate the meaning of the "level" and its utility during the analogical exploration.

3.1 Example 1: Component-, Behavior-, and Material-Based Analogies for Pocket Tool.

The component result shows patent pocket tool and patent "electric toothbrush" in a cluster label of "switch, circuit, battery, house, port," suggesting that they are composed of small electronic components. The behavior result shows the same patent pocket tool and patent "method for protecting electric line" in a cluster label of "electronic, electric, peripheral, mechanic, secure," suggesting that they have a commonality of protecting and securing electronic components. Lastly, the material result shows the same patent "pocket tool" and patent "deburring knife with replaceable blade" in a cluster label of "arrow, metal, solid, waste, stem," suggesting that they are either made of metal or use metal.

3.2 Example 2: Component- and Behavior-Based Analogies for Ice Gripping Sandal.

In a second example, the component result shows patent ice gripping sandal and patents "weight distributing knee pad" and "balance assist for rotating recreational device" in a cluster label of "port, strap, case, mount, heel." The behavior result shows the same patent ice gripping sandal and patents "abrasive tool" and "hairpiece and fitting method therefor" in a cluster label of "secure, flex, light, sole, alternative." Although the cluster labels of the component and behavior results are different, the patents, interestingly, share a common functionality of securing or fixing something to something else. This is an example of discovering patents of similar design attributes from various apparel domains using component and behavior data. For instance, patent ice gripping sandal is from the footwear domain, weight distributing knee pad is from wearing apparel domain, and hairpiece and fitting method therefor" is from headwear domain.

3.3 Example 3: Component- and Material-Based Analogies for Shower Bath Apparatus and Spray Nozzle.

In a third example, the component result shows patent shower bath apparatus and spray nozzle and patent "transmucosal hormone delivery system" in a cluster label of "nozzle, spray, bath, body, valve," suggesting that they use certain components to deliver fluids. The material result shows the same patent shower bath apparatus and spray nozzle and patent "humidifier" in a cluster label of "water, fluid,

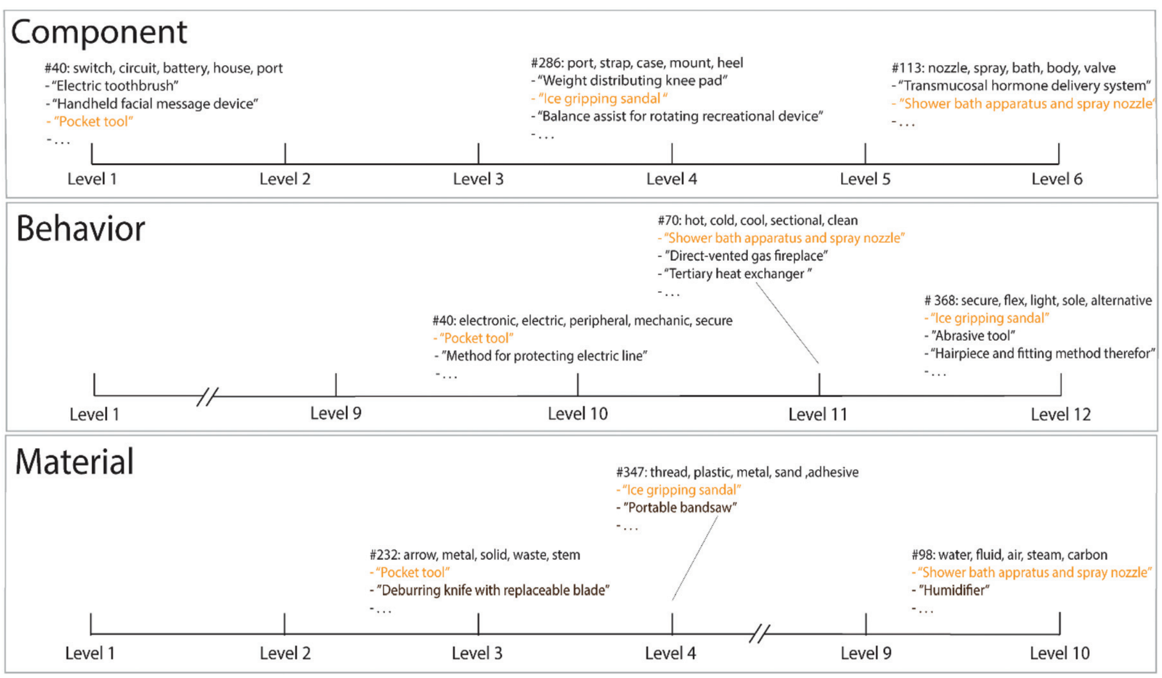

Fig. 4 Comparison of component, behavior, and material results

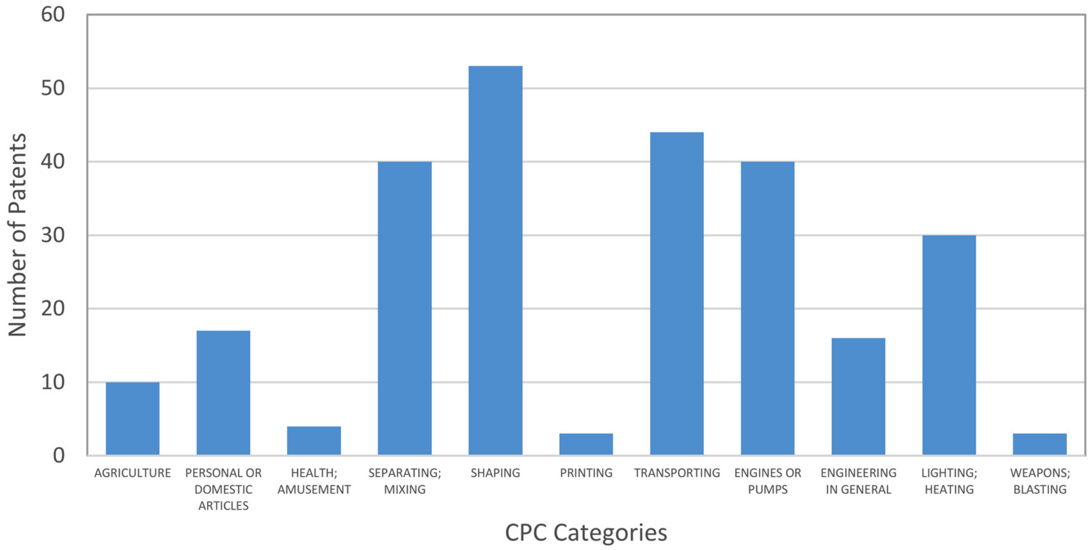

Fig. 5 Number of patents in different CPC categories

Table 1 Retrieved patents in different CPC sections

Cluster number (cluster label)	Patent title (patent number)	CPC section
134 (shoe, mount, boot, heel, generator)	"Inflatable boot liner with electrical generator and heater" (4845338)	A43 (Footwear)
123 (motor, generator, shaft, detector, crankshaft)	"Gravity motor and method" (7768142)	F3 (Machines or engines for liquids)
121 (disk, shaft, wheel, ramp, transmission)	"Continuously variable transmission" (7147586)	B62 (Land vehicles for travelling otherwise than on rails)

air, steam, carbon," suggesting that water or fluid is the common material used by the two designs. Although the two cluster labels represent different analogical relationships, they all have a common functionality of delivering fluids—one that is implied by the components such as "nozzle" and "spray" and one that is implied by the liquid material that the design uses.

As shown in the examples, exploring different representations of the patent data leads to retrieving multiple analogies that can be valuable for design-by-analogy process. In practice, a designer may explore the multiple representations to retrieve more analogies to improve the idea generation [8]. Also, they may explore a specific representation of the patents that is better matched with their individual perspective or way of thinking.

4 Case Study and Discussion

To test the usefulness of the generated patent structures to design-by-analogy practice, a case study has been conducted to examine the retrieval of analogies for idea generation. The case study is hypothetical use case scenario, in which one of the researchers used the computational methodology for idea generation. The case study introduces two patent exploration methodologies, referred to as (1) searching similar patents in different CPC sections and (2) searching patents by suggested cluster label. This section discusses the step-by-step procedures of retrieving analogies from the component- and behavior-based patent structures to solve the human motion energy collection problem introduced in Sec. 2.1.

4.1 Method 1: Searching Similar Patents in Different Cooperative Patent Classification Sections Using Component Data. For the given design problem shown in Sec. 2.1, a designer first brainstorms relevant terms for generating electrical energy from human motion. For example, if the designer wants to explore patents by their components, a list of terms might be "generator," "turbine," "wheel," "shaft," "gear," "rotor," "tooth," "pulley," "disk," and "crankshaft." These component terms are selected for their rotational movement, which is the most common type of mechanical energy that is converted to electrical energy. The designer then uses the list to select patents whose cluster label matches the component terms. As discussed in Sec. 2.2, the cluster label is composed of the top five most important words to the group of patents in the cluster, determined by NMF, that characterize the cluster's topic. In such a case, the large patent dataset is refined to a few clustered patents that are operated by or composed of the components that involve rotational energy/movement. After the patent data are reduced, the designer categorizes the patents by their CPC section. Categorizing the patents by their CPC sections not only helps the designer to understand patent's usage and application, but also helps the designer to recognize similar technology in various domains. Consequently, this method results in a reduced dataset of 193 patents in 11 different fields of application as shown in Fig. 5.

The number of retrieved patents and analogies are expected to increase with an increased input patent corpus size, which is currently set to 1060 patents in this study. The designer then explores the refined patent data by the field of application. Table 1 lists three patents from machine, footwear, and land vehicle domains that have useful analogies to solve the design problem. For example, one resulting design concept is an inflatable floor mat (inspired by patent "inflatable boot liner with electrical generator and heater") that uses human body weight to pump and direct air across an electrical generator. Furthermore, the designer can implement a gravity motor (inspired by patent "gravity motor and method") that uses gravitational energy to rotate a shaft or continuously variable transmission (inspired by patent "continuously variable transmission") that improves the shaft's torque efficiency. The concepts generated are shown in Fig. 6.

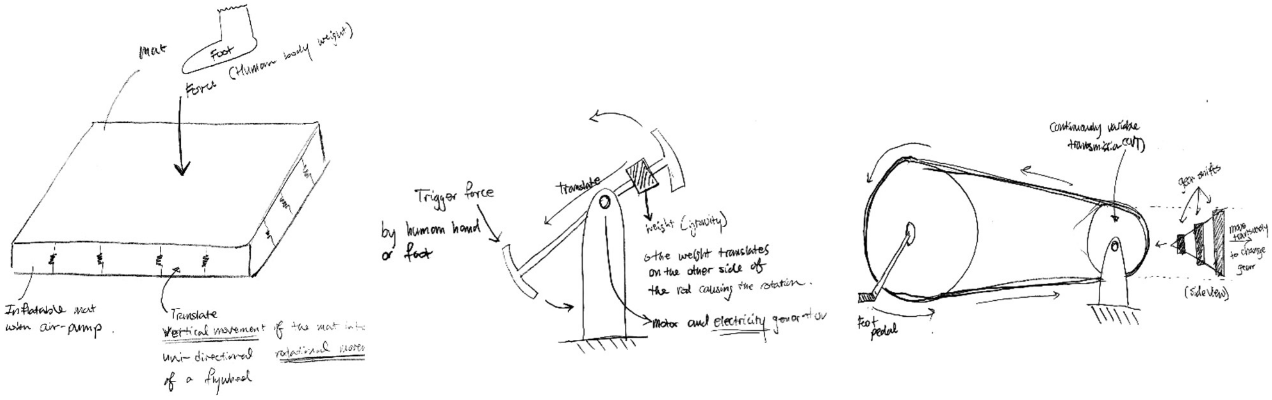

Fig. 6 Concepts generated using method 1

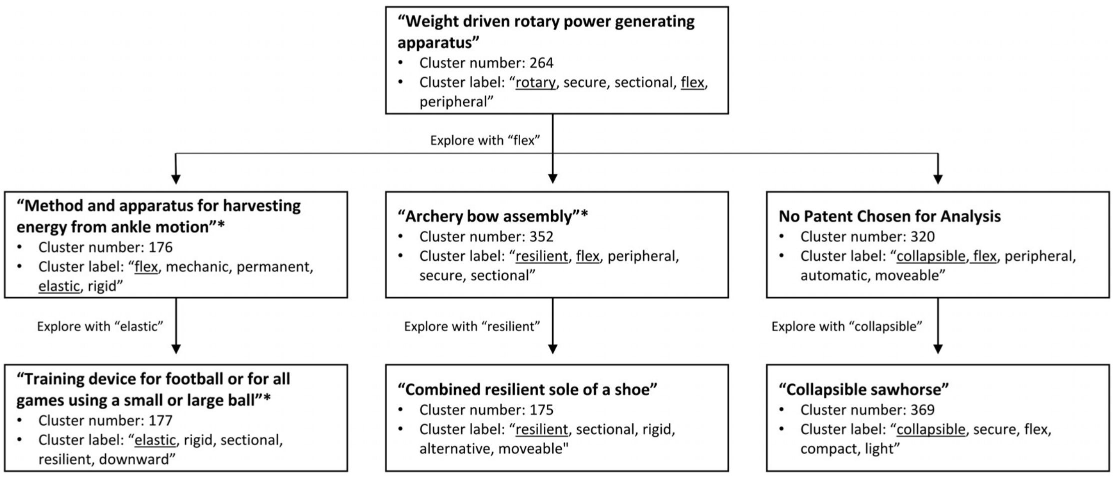

Fig. 7 Patents (bolded) retrieved using suggested terms (underlined)

4.2 Method 2: Searching Patents by Suggested Terms in Cluster Label Using Behavior Data. For the same design problem, an alternative method may be used to retrieve more analogies from the patent data. Similar to the first method, the designer first generates a list of relevant terms for the design problem. For example, if the designer wants to explore the behavior-based patent structure, a set of behavior terms would be "rotational" and "rotary" as the rotational movement is conventionally used for generating electricity. The designer then explores the patents by matching their cluster labels to the behavior terms. The cluster label consists of the top five words that are closely related to each other by an NMF-determined topic. Figure 7 demonstrates that the designer first retrieves a patent in a cluster "rotary, secure, sectional, flex, peripheral." In this label, the designer discovers another term "flex," implying that some patents may integrate both rotational and spring behaviors. Using the new keyword, "flex," the designer retrieves two more patents and discovers three more keywords "elastic," "resilient," and "collapsible" that result in retrieving three more patents. Interestingly, in Fig. 7, patents with an asterisk mark (*) after the title represent patents with combined elastic and rotational behaviors. This exemplifies the retrieval of analogies using distinct yet suggested terms in the cluster label, with an emphasis on exploration and discovery. Figure 8 shows three example design concepts generated using patents "method and apparatus for harvesting energy from ankle motion," "training device for football or for all games using a small or large ball" and "archery bow assembly."

The two patent searching methods can be used interchangeably for an effective retrieval of analogies. The first method demonstrates that if a list of keywords for the desired technology is available or can be generated by a designer, a patent structure can be refined with the keywords and further refined by CPC sections. The refining or categorizing of the patent data can enable designers to systematically and efficiently retrieve similar analogies in various application fields. If the list of keywords is not available, the second method demonstrates that the patents can be iteratively searched with the suggested terms in the cluster labels. This approach is more spontaneous as designers can determine the exploration paths using the cluster label while interacting with the patent structure.

As demonstrated by the two methodologies, the patent exploration depends significantly on the usage of keyword search, as it greatly reduces the analogical information that designers need to process. Moreover, the user-defined keywords allow the designers to tailor the patent exploration to focus on specific analogies that satisfy the designers' objectives for a design problem. In the case study, the component and behavior terms relating to the rotational movement/energy are used to solve the human motion energy collection problem. As discussed earlier, designers may have diversified design objectives, implying that the list of the terms is just one of many paths to approach the design problem. These motivations call for future studies to generate lists of words that address specific working or physical principles of patents and investigate the effectiveness of refining the patent structure using the pre-established list of words.

4.3 Clustering Quality of Material Data. In the case study, the material-based patent structure has not been studied due to the difficulty of identifying and retrieving analogies. Unlike the component and behavior results, the cluster labels in the material result do not carry definite topics. For instance, terms like "water," "fluid," "air," "gas," and others appear in many cluster labels, causing the clusters to be indistinguishable from one another. To

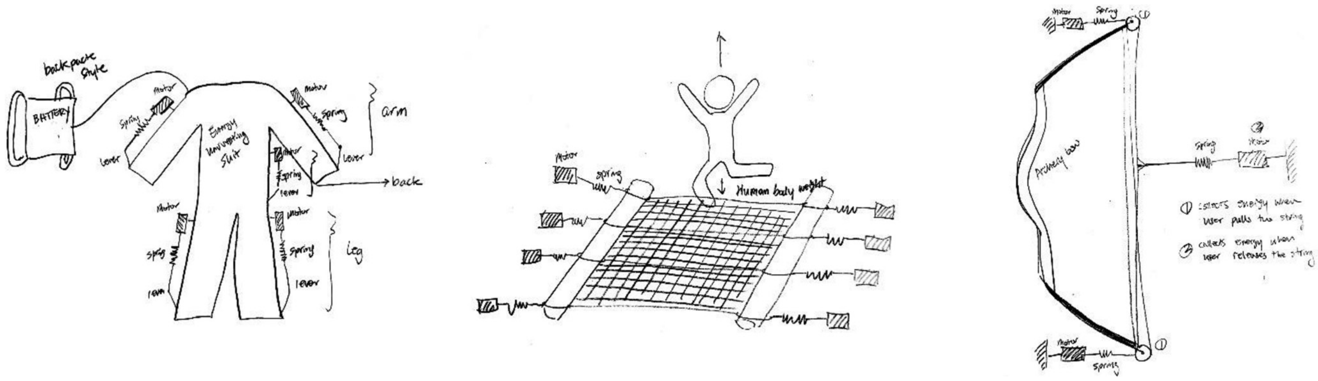

Fig. 8 Concepts generated using method 2

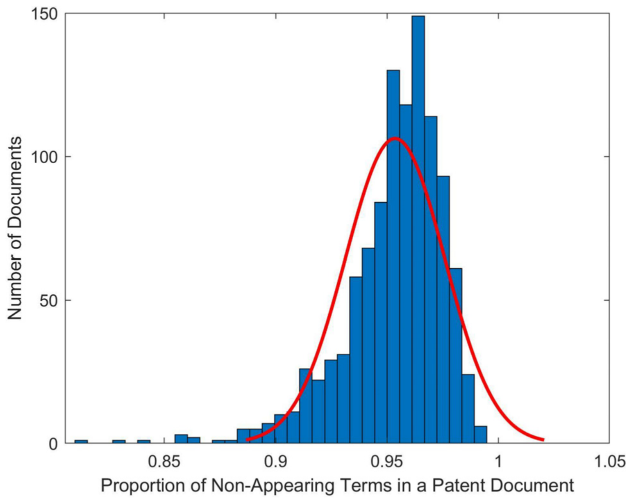

Fig. 9 Histogram of component result

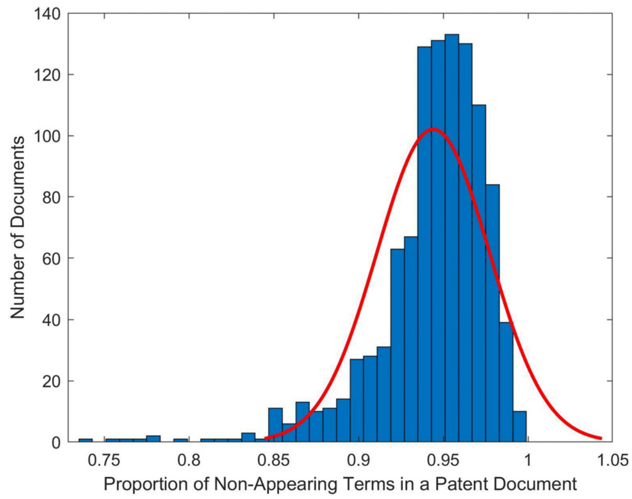

Fig. 10 Histogram of behavior result

understand why the cluster quality is relatively poor for the material data, the word-document matrices of component, behavior, and material data are studied. Figures 9–11 show the histograms of three different datasets that represent the distribution of the number of patent documents (y-axis) by the proportion of terms that *do not* appear in each document (x-axis). The proportion value varies from 0 (the document contains all terms listed in a wordlist) to 1 (the document does not contain any of the terms listed in a wordlist). The histograms show a general bell curve for all datasets, where the mean values are 0.954 for component data, 0.944 for behavior data, and 0.966 for material data. Although the mean values are similar, it is clearly observed that the histogram for the material data is skewed to the right. The skewed data represent approximately 500 patents that contain less than 3% of the material terms and approximately 150 patents that contain zero of the material terms. This finding explains the poor cluster quality of the material data, as NMF would randomly determine the topic of the patent documents whose column cells are empty. As discussed in Sec. 2.1, the wordlists for the component, behavior, and material properties in Appendices C, D and E were compiled by drawing a master wordlist from the entire patent data and manually categorizing the words into an appropriate property. This ensures that the compiled material wordlist captures all material words of the input patent data. In such a case, the absence of material words in many of the patent documents raises some interesting research questions about the nature of patent documents: to what extent do patent documents describe the materials that the system/artifact/technology use? When documenting for patents, do designers or inventors detail the functional features of the system/artifact/technology to protect the functional quality of the intellectual property, but omit what the input or comprising

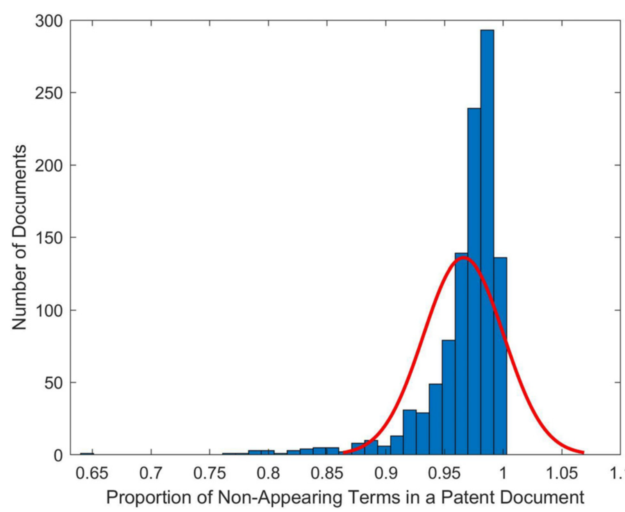

Fig. 11 Histogram of material result

materials in the system/artifact/technology to ensure the broadest protection from infringement? Future studies need to investigate whether certain types of words are used more frequently than others for documenting patents. Interviews with patent lawyers on their writing process and choices may help to uncover some of the driving factors behind the composition of the semantic data. Understanding the unique traits of patent documents could help researchers to improve the quality of topic modeling of patent data.

4.4 Limitations and Future Work. There are several limitations in the proposed methodology that are important to acknowledge. It has been observed that the clustering quality is poor when there is a substantial number of patents that contain ambiguous or indeterminable topics. This suggests that the clustering performance of the current methodology may not be competent for a larger set of patents. It is also uncertain that the current visualization methods are appropriate for interpreting the larger patent data. Collaborations with computer scientists on investigating clustering and visualization methods may lead to developing more effective DbA tools for designers. The case study shows that the material-based clustering is not useful in assisting the design-by-analogy process. This may be a product of the dataset examined in the study or a more widespread phenomenon, such as how patents are documented, that needs to be investigated in future studies. In addition, the wordlists that are manually compiled may not capture all relevant content in the patent data. However, we expect the content of these lists to reach saturation as larger data sets are tested in future studies. Other limitations and future work have been discussed throughout the paper including, but not limited to, the implementation of improved stopping criterion for the iterative NMF process to generate the clustering structures and investigation of the meaning of the "level of similarity" for its utility during analogical exploration process.

5 Conclusion

The goal of this work is to provide designers a computational design-by-analogy tool to explore a design repository for analogical inspiration. The computational methodology utilizes a topic modeling technique called NMF to generate a hierarchical cluster structure of the U.S. patent database where designers can visualize and retrieve analogies to solve given design problems. In this work, a user-driven exploration technique has been implemented on the patent database, processed to contain only component-, behavior-, or material-based content to generate a unique and valuable representation of the patent database; a case study was conducted to demonstrate the systematic and interactive retrieval of analogical inspirations from patents for concept generation. Through the demonstration of retrieving analogical inspiration from the structured patent database for concept generation, the proposed methodology shows promise as an effective design-by-analogy tool that facilitates the systematic and interactive exploration of patents to support innovative design idea generation.

Acknowledgment

This work is supported by the National Science Foundation, under grant CMMI 1663204. The United States Government retains, and by accepting the article for publication, the publisher acknowledges that the United States Government retains, a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for United States Government purposes.

Funding Data

- National Science Foundation (Grant No. CMMI 1663204; Funder ID: 10.13039/501100008982)

Appendix A

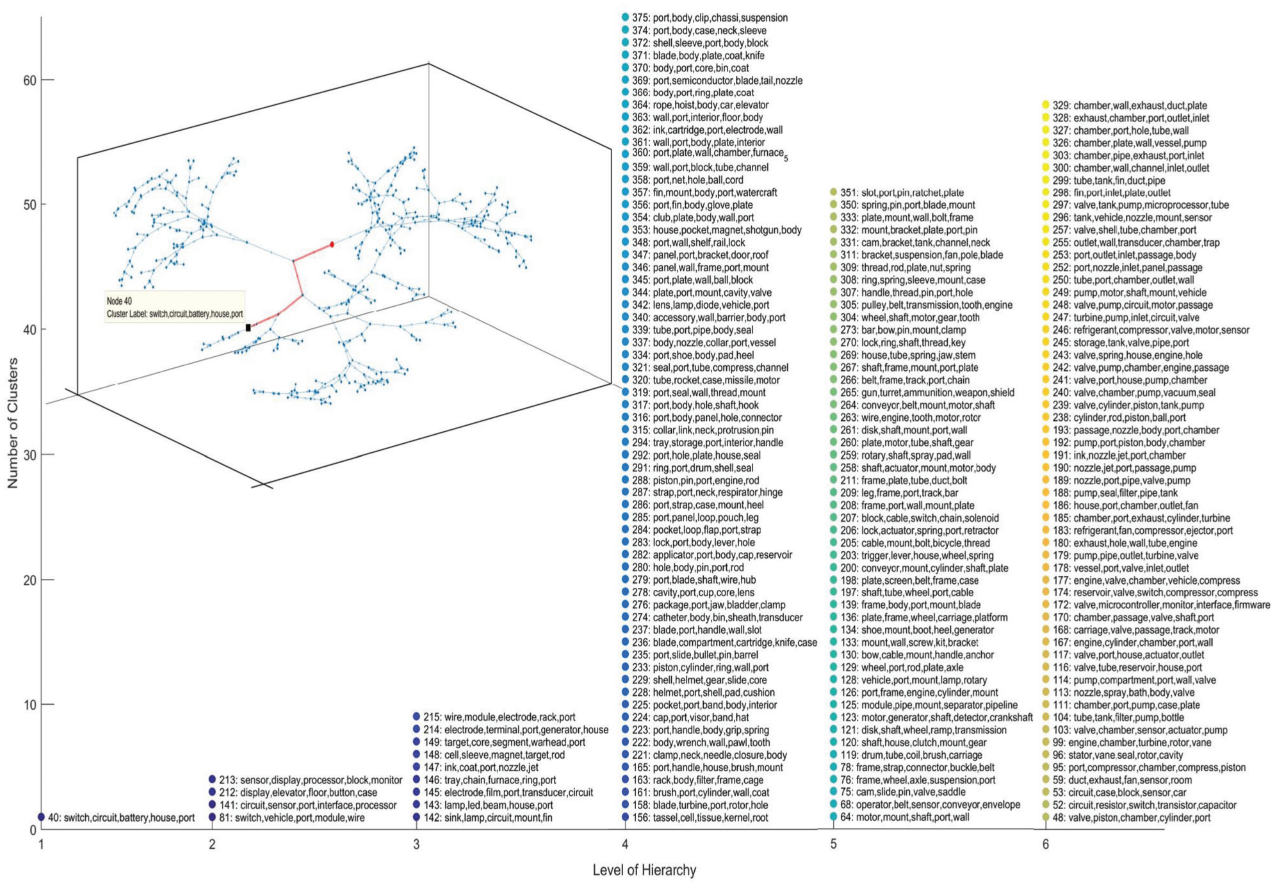

Fig. 12 Three-Dimensional and Two-Dimensional Visualizations for Component Representation of Patent Data

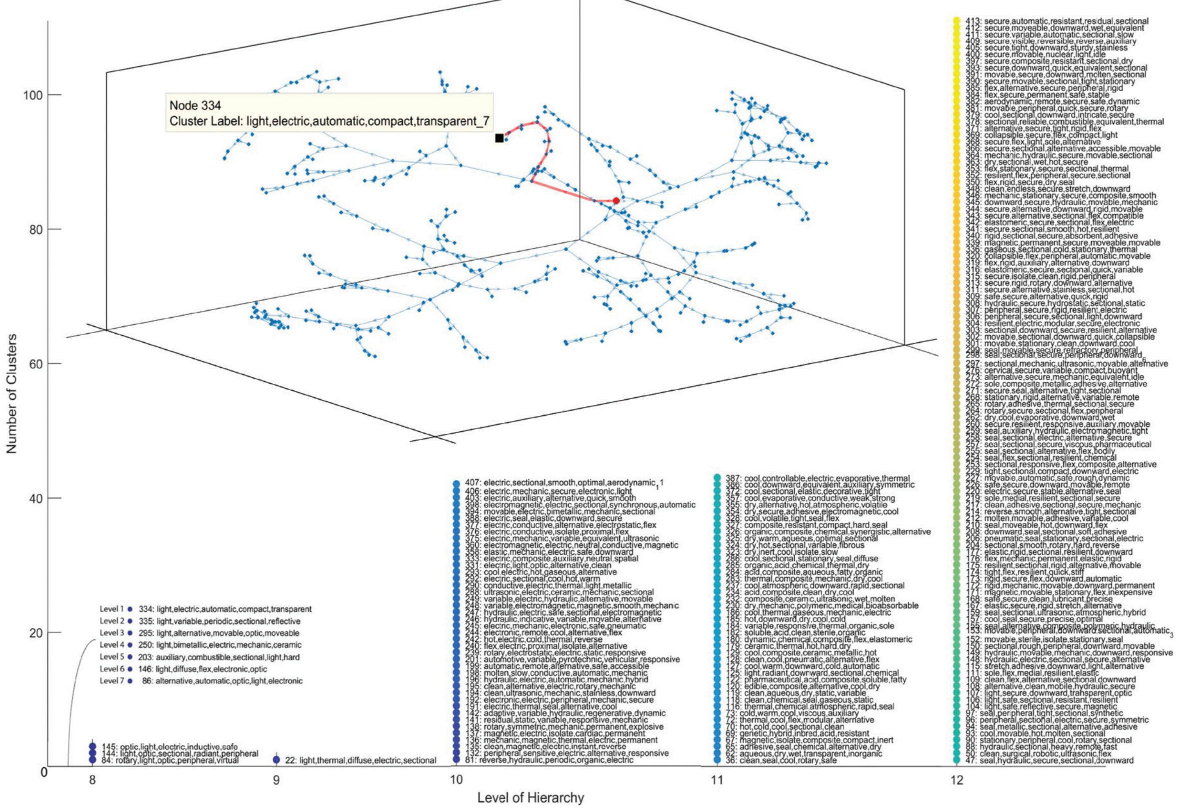

Fig. 13 Three-Dimensional and Two-Dimensional Visualizations for Behavior Representation of Patent Data

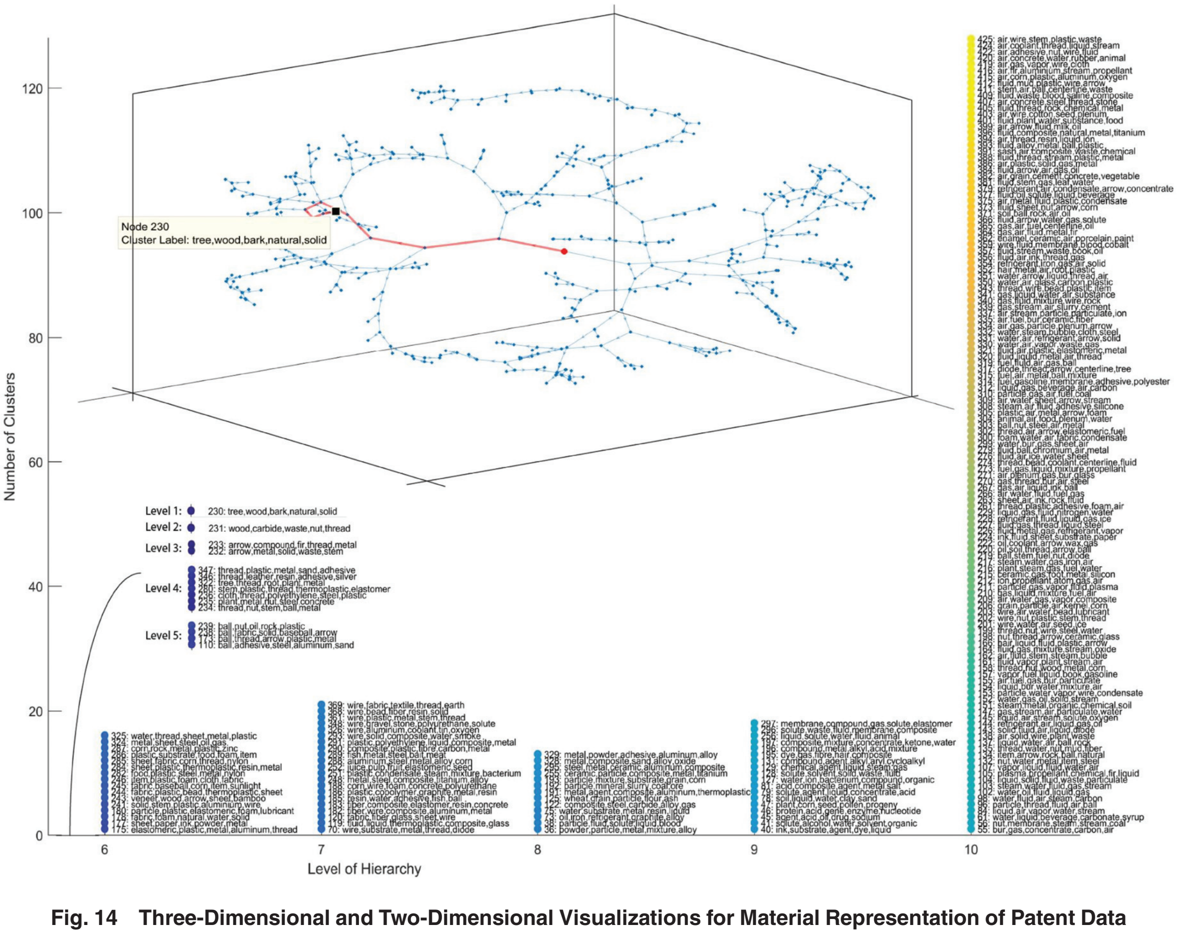

Fig. 14 Three-Dimensional and Two-Dimensional Visualizations for Material Representation of Patent Data

Appendix B: List of 53 Cooperative Patent Classification Subsections

Section	Subsection	Categories	Description
Section A: Human necessities	1	Agriculture	Agriculture; forestry; animal husbandry; hunting; trapping; fishing
	41	Personal or domestic articles	Wearing apparel
	42		Headwear
	43		Footwear
	45		Hand or traveling articles
	46		Brushware
	47		Tables; desks; office furniture; cabinets; drawers; general details of furniture
	61	Health; amusement	Medical or veterinary science; hygiene
	62		Life-saving; fire-fighting
	63		Sports; games; amusements
Section B: Performing operation; transporting	2	Separating; mixing	Crushing, pulverizing, or disintegrating; preparatory treatment of grain for milling
	3		Separation of solid materials using liquids or using pneumatic tables or jigs; magnetic or electrostatic separation of solid materials from solid materials from solid materials or fluids; separation by high-voltage electric fields
	5		Spraying or atomizing in general; applying liquids or other fluent materials to surfaces, in general
	6		Generating or transmitting mechanical vibrations in general
	7		Separating solids from solids; sorting
	8		Cleaning
	9		Disposal of solid waste; reclamation of contaminated soil
	21	Shaping	mechanical metal-working without essentially removing material; punching metal
	22		Casting; powder metallurgy
	23		Machine tools; metal-working not otherwise provided for
	24		Grinding; polishing
	25		Hand tools; portable power-driven tools; manipulators
	26		Hand cutting tools; cutting; severing
	27		Working or preserving wood or similar material; nailing or stapling machines in general
	28		Working cement, clay, or stone
	29		Working of plastics; working of substances in a plastic state, in general
	41	Printing	Printing; lining machines; typewriters; stamps
	60	Transporting	Vehicles in general
	61		Railways
	62		Land vehicles for traveling otherwise than on rails
	63		Ships or other waterborne vessels; related equipment
	64		Aircraft; aviation; cosmonautics
	65		Conveying; packing; storing; handling thin or filamentary material
	66		Hoisting; lifting; hauling
	67		Opening, closing (or cleaning) bottles, jars, or similar containers; liquid handling
Section F: Mechanical engineering; lighting; heating; weapons; blasting	1	Engines or pumps	Machines or engines in general
	2		Combustion engine
	3		Machine or engines for liquids
	4		Positive displacement machine for liquids; pumps for liquids or elastic fluids
	5		Indexing schemes relating to engines or pumps in various sub-classes of classes
	15	Engineering in general	Fluid-pressure actuators ; hydraulic or pneumatics in general
	16		Engineering elements and units; general measures for producing and maintaining effective functioning of machines or installations; thermal insulation in general
	17		Storing of distributing gases or liquids
	21	Lighting; heating	Lighting
	22		Steam generation
	23		Combustion apparatus
	24		Heating; ranges; ventilating
	25		Refrigeration or cooling; combined heating and refrigeration systems; heat pump systems; manufacture or storage of ice; liquefaction solidification of gases
	26		Drying
	27		Furnaces; kilns; ovens; retorts
	28		Heat exchange in general
	41	Weapons; blasting	Weapons
	42		Ammunition; blasting

Appendix C: Wordlist of Components

abaci
abacus
absorber
accelerator
accelerometer
accessory
accumulator
activator
actuation
actuator
adapter
adaptor
adjuster
adjustor
afterburner
agitator
ailette
airfoil
airgun
airplane
alarm
alignment
alternator
ammunition
amplifier
analyzer
anchor
anchorage
annulus
annunciator
antenna
anticipator
applicator
aquarium
armature
armrest
arrow
arrowhead
artery
aspirator
atomiser
atomizer
atrium
attemperator
auditorium
autoinjector
automobile
ax
axle
backbone
backpack
backpack
backrest
badge
bag
bagatelle
ball
balloon
band
bandage
bar
barbell
barcode
barometer
barrel
barrier
base
basin
bath
bathseat
bathtub
battery
beam
beanbag
bear
bearing
bed
bedstead
bell
bellows
belt
bench
bicycle
bike
billard
bin
bingo
biomarker
birdhouse
bladder
blade
block
board
boat
body
bogie
boiler
bolt
bone
booster
boot
border
bottle
bottles
bow
bowl
bowstring
box
brace
bracelet
bracket
brake
branch
breadboard
breech
brewer
bridge
bronchoscope
brooch
brush
bucket
buckle
bulb
bullet
bundle
bus
button
buzzer
cabinet
cable
cage
calculator
calibrator
caliper
cam
camera
camshaft
candle
canister
cannister
cannula
canoe
canopy
cantilever
cap
capacitor
capsule
car
carabiner
carbine
carburetor
carburettor
cardia
carousel
carpet
carriage
carrier
cart
cartridge
cartridges
case
cassette
castor
catalyst
catapult
catheter
cathetor
catridge
cavaletti
cavity
ceiling
cell
centimeter
centrifuge
chain
chair
chalk
chamber
channel
chassi
chimney
chip
circuit
clamp
classifier
cleaner
clip
clock
closure
cloth
club
clutch
clutches
coat
cogwheel
coil
colander
collar
collector
collimator
combustor
commutator
compactor
comparator
compartment
compensator
component
compress
compressor
computers
condenser
condensor
conditioner
conditioners
conductor
connecter
connector
conrod
console
contactor
container
containers
contractor
convertor
conveyor
cooker
cookstove
cooktop
coordinator
cord
core
cork
cornea
corset
cotter
countermeasure
countershaft
cover
cradle
crankcase
crankshaft
cribbage
crinoline
crossbar
crossbow
crown
cuff
cufflink
cup
cursor
curtain
cushion
cuspidor
cylinder
cytometer
dam
dampener
damper
dashboard
dashpot
deck
decoder
defibrillator
deflector
defroster
dehydrator
demodulator
descrambler
descriptor
designator
desk
desktop
desuperheater
detector
detonator
dial
dialyzer
diaphragm
diffuser
dilator
diode
disk
discriminator
dish
dishwasher
disk
dispenser
displacer
display
disruptor
distractor
distributor
divider
dome
door
doublet
dowel
drawstring
drier
drill
drum
duct
dumbbell
durometer
dynamometer
economiser
economizer
editor
effector
ejector
electrode
electroluminescent
electromagnet
electronics
elevator

eliminator
emitter
emulsifier
encoder
endoscope
engine
enhancer
envelope
equalizer
equalizer
equipment
estimator
evaluator
evaporator
examiner
excavator
exciter
exhaust
expander
extender
extension
extinguisher
extruder
eyelet
eyepiece
fan
fastener
fastening
femur
fence
fermenter
fermentor
ferrule
fertilizer
festoon
filament
filler
film
filter
fin
fins
firearm
firebridge
fireplace
firmware
flameholder
flap
flash
flashlight
flask
floor
flue
flute
flywheel
foil
footwear
forceps
forearm
fork
frame
freewheel
furnace
fuze
gaiter
gallery
gamepad
garland
garter
gasket
gastroscope
gate
gauge
gear
gearbox
gears
gearset
generator
geometrizer
girdle
glove
glowplug
glue
goggle
gown
graft
granulator
grease
grenade
grid
grip
guidewire
guitar
gum
gun
handle
handled
hardware
harness
harpoon
hat
headband
headgear
headrest
headset
heart
heel
helicopter
helmet
highway
hinge
hoist
holder
hole
holster
homogenizer
honeycomb
hood
hook
hoop
hopper
horn
horseshoe
hose
hosel
house
housing
howitzer
hub
hypotube
identifier
idler
igniter
ignitor
illuminator
impeller
inclinometer
incubator
indicator
inducer
inductor
inflator
inhaler
inhibitor
initiator
injector
injector
ink
inlet
installer
instance
insulator
integrator
interconnector
intercooler
interface
interior
interrupter
intestine
introducer
investigator
isolator
jack
jacket
jaw
jet
jets
joint
joypad
joystick
kernel
key
keyboard
keypad
kit
kite
knife
knob
lamp
lanyard
laser
lash
latch
layer
leaflet
leash
led
leg
legging
legs
lens
lenslet
lever
lid
lift
lighter
liner
lining
link
linkage
lip
liquefin
lobe
localizer
locator
lock
lodge
longbow
loop
loudspeaker
lubricant
luggage
luminaire
magnet
maintainer
mandrel
manifold
manipulator
manometer
mantle
marker
mask
mast
mat
mattress
mediator
membrane
microcomputer
microcontroller
microlens
micrometer
microphone
microprocessor
microscope
microswitch
midline
mirror
missile
modifier
modulator
module
moisturizer
monitor
motherboard
motor
motorcycle
mouldboard
mount
mounted
mousetrap
mouthpiece
mover
mower
muffler
multicylinder
multiplate
multiplier
multivibrator
muscle
muscles
muzzle
nanometer
neck
neckband
needle
nerve
net
neurotransmitter
nostril
notebook
notepad
nozzle
nut
observer
obturator
occluder
opacifiers
operator
ordnance
organ
orifice
oscillator
oscilloscope
outlet
oven
oversleeve
oxidizer
oximeter
pacemaker
package
pad
paddle
paintball
pan
pane
panel
panel
parachute
parameter
parlor
passage
passageway
patch
path
pawl
pedal
pedestal
peg
pen
pencil
penetrator
penis
peppergun
perforation
petticoat
phone
photodetector
photosensitizer
pillar
pillow
pin
pinion
pipe
pipeline
pistol
piston
pitot
planetary
plasticizer
plate
platelet
platform
player
playhouse
playslip
pleximeter
plug
pocket
pod
pointer
polarizer
pole
pool
pore
port
portafilter
potentiometer
pouch
powertrain
prechamber
precipitator

precursor
predictor
prefilter
pressurizer
preventer
probe
processor
progenitor
projector
promoter
propeller
protector
protrusion
provider
puck
pulley
pump
puttee
puzzle
rack
racket
radar
radiator
radio
rail
railroad
ramp
ratchet
razor
reactor
reboiler
receiver
receptor
rectifier
reducer
reel
reflector
refrigerant
refrigerator
regenerator
regulator
repeater
resectoscope
reservoir
resistor
respirator
restorer
restricter
restrictor
resuscitator
retainer
retina
retractor
rib
ribbon
riblet
ribs
rifle
riflescope
rille
rim
ring
rips
rivet
robot
rocket
rod
rodlink
roller
roof
room
root
rope
rotary
rotator
rotor
sack
saddle
sandbox
satellite
scaffold
scale
scattergun
scene
scheme
scissors
scooter
scope
screen
screw
screwdriver
scroll
seal
seat
sector
see-saw
segment
segments
selector
semiconductor
sensor
separator
servomotor
sewer
shack
shackle
shaft
shank
shaver
sheath
shelf
shell
shield
ship
shoe
shoelace
shooter
shotgun
shredder
shuttle
sidewall
sieve
silencer
simulator
sink
siren
ski
skin
skirt
sleeve
slide
sling
slot
slug
sluice
snare
sock
socket
socles
solenoid
solubilizer
speaker
spear
speargun
spectrometer
sphincter
spigot
spindle
spine
spinner
spittoon
spool
spoon
spray
spring
sprocket
stabilizer
stand
staple
station
stator
stem
stent
stents
stethoscope
stiffener
stile
stimulator
stock
stool
storage
stove
straightener
strap
stratifier
straw
streamline
string
stroke
stud
subchamber
subsoiler
suction
sulky
supercharger
suppressor
suspension
suture
swashplate
swing
switch
swivel
synchronizer
synthesizer
syringe
table
tablet
tack
tail
tampon
tank
tape
tappet
target
tassel
teaspoon
teeth
telemetry
telemotor
telephone
television
tendon
terminal
tether
thermister
thermistor
thermocouple
thermometer
thermosensor
thermoset
thermostat
thorax
thread
throttle
thyristor
ticket
tie
tile
timer
tip
tire
tissue
tongue
tooth
toothbrush
torch
toupée
towel
toy
tracheal
track
trackball
tractor
trail
transceivera
transducer
transformer
transistor
transmission
transmitter
transometer
transponder
trap
tray
trebuchet
trigger
trimmer
trocar
trolley
trommel
trough
truck
trunk
tub
tube
tubes
tunnel
turbine
turbo
turbocharger
turboexpander
turbofan
turbomachine
turbomachinery
turbopump
turf
turnbuckle
turntable
turret
ultrafilter
umbrella
ureter
vacuum
valance
valve
vane
vaporizer
vault
vehicle
vein
velcro
vent
ventilator
venturi
vessel
vibrator
visor
wagon
wall
wand
warhead
washer
watch
watercraft
weapon
wedge
wheel
wheelchair
whiteout
window
wing
wire
wrench
yoke
zipper

Appendix D: Wordlist of Behaviors

abdominal
absorbable
absorbent
accessible
acetabular
acid
acidic
actuatable
actuated
adaptable
adaptive
adherable
adhesive
adverse
aerodynamic
agricultural
alignable
alternative
anatomic
anhydrous
aqueous
aromatic
arterial
artificial
asymmetric
atmospheric
attachable
audible
automated
automatic
automotive
auxiliary
bacterial
ballistic
bendable
bimetallic
bioabsorbable
biocompatible
biodegrade
biologic
bipolar
bloated
bodily
breathable
buoyant
cardiac
cellular
ceramic
cervical
chemical
chronic
clean
clinic
coated
cold
collapsible
combustible
compact
compatible
complementary
complex
composite
compressed
compressive
condensed
conductive
conjoint
consistent
controllable
controlled
convective
cool
coronary
cosmetic
cryogenic
curable
decorative
deflectable
deformable
demountable
dense
dental
deployable
detachable
detectable
diffuse
discrete
displaceable
disposable
dissolvable
distal
divisional
downward
dry
durable
dynamic
edible
elastic
elastomeric
electric
electromagnetic
electronic
electrostatic
electrosurgical
elevated
elevational
endless
engageable
enlarged
environmental
equivalent
evaporative
exchangeable
expandable
expanded
explosive
extendable
extended
extendible
extracellular
fast
fastenable
fatty
femoral
fetal
fibrous
fitting
flammable
flex
flexural
floatable
floating
floral
fluidic
fluorescent
fragile
fresh
frictional
fungicidal
fusible
gaseous
gastrointestinal
genetic
hard
harmonic
healthy
heavy
hemispheric
hemostatic
hot
hybrid
hydraulic
hydrodynamic
hydrophilic
hydrophobic
hydrostatic
hypodermic
idle
ignitable
immovable
immune
impermeable
implantable
inbred
incompressible
indicative
inducible
inductive
inelastic
inert
inertial
inexpensive
inflammatory
inflexible
infrared
injectable
inorganic
insertable
insoluble
instant
instantaneous
insulative
interchangeable
interconnecting
intravenous
intricate
invasive
inverted
invertible
irregular
irreversible
isolate
latent
light
lubricant
lumbar
luminous
magnetic
magnetizable
malleable
mechanic
medial
medical
metallic
miscible
mixed
mobile
modular
molten
mountable
movable
moveable
muscular
myocardial
nasal
nestable
neutral
noncombustible
nonionic
nonrotatable
nuclear
nucleic
ocular
opposing
optic
optimal
orbital
organic
periodic
peripheral
permanent
permeable
pharmaceutical
pharmacologic
physical
piercing
pivotable
pivotal
planetary
pneumatic
polymeric
positionable
precise
premature
programmable
programmed
prolonged
propelled
proximal
pure
pyrotechnic
quick
radiant
radioactive
rapid
rechargeable
recombinant
recyclable
reflective
refractory
regenerative
releasable
reliable
remote
repellent
repetitive
replaceable
reproducible
residual
resilient
resistant
resonant
responsive
resting
reusable
reverse
reversible
revolving
rigid
robotic
rollable
rolling
rotary
rotatable
rough
rugged
safe
saline
saturated
seal
sectional
securable
secure
semipermeable
sensible
sensitive
shiftable
shrinkable
slidable
slideable
sliding
slow
smooth
soft
solar
sole
soluble
spatial
spinal
spinning
spinous
stabilized
stable
stackable
stainless
static
stationary
steady
steerable
sterile
sticky
stiff
stretch
strong
sturdy
submersible
supporting
surgical
susceptible
sustained
swivelable
symmetric
synchronous
synergistic
synthetic
systolic
tangential
therapeutic
thermal
thermodynamic
tight
tiltable
torsional
toxic
transluminal
transparent
transportable
turbulent
ultrasonic
underwater
unstable
variable
vascular
vehicular
ventricular
vertebral
virtual
viscous
visible
volatile
warm
waterproof
waxy
weak
wearable
wet
wheeled

Appendix E: Wordlist of Materials

acetate
acid
acridine
adhesive
adsorbent
agent
air
albumin
alcohol
algae
alkali
alkaline
alkaloid
alkenyl
alkoxy
alkyl
alkylcarboxamide
alkylene
allantoin
alloy
alum
alumina
aluminum
aluminum
amide
amine
aminoalkyl
ammonium
animal
antibacterial
antibiotic
antibody
arrhythmia
arrow
arthropod
aryl
ash
asphalt
atom
bacterium
bait
ball
bamboo
banana
bark
baseball
bauxite
bead
bean
beeswax
benzotriazole
benzyl
beryllium
beverage
biodiesel
biopolymer
blood
book
borax
brass
brick
bromide
bronze
bubble
bur
calcium
cancer
canvas
carbide
carbohydrate
carbon
carbonate
carboxyl
cast iron
castor
cellulose
cement
ceramic
chalk
charcoal
chemical
chloride
chlorine
chocolate
chromium
claudin
clay
cloth
coal
cobalt
coffee
collagen
composite
compost
compound
concentrate
concrete
condensate
coolant
copolymer
copper
corn
cotton
crab
cream
crystal
curd
cycloalkyl
daub
dextran
dialysate
diamond
diesel
dimethyl
diode
dioxide
disulfide
dough
drug
duroplastic
dye
dynamite
earth
egg
elastomer
elastomeric
electrolyte
enamel
enantiomer
endocardium
enzyme
epoxy
espresso
ester
ethanol
ether
ethyl
ethylene
fabric
fat
feedstuff
fenugreek
fiber
fiberglass
fiber
fir
fish
flour
fluid
fluoride
fluorine
fluorite
foam
food
foodstuff
fructose
fruit
fuel
fungicide
galactosyl
gas
gasoline
gel
gem
gemstone
gene
glass
glucose
glyceryl
glycol
gold
grain
granite
graphite
gravel
hair
halogen
hay
hemoglobin
hemp
herbal
herbicide
heteroaryl
honey
honeycomb
humectant
huperzine
hydride
hydrocarbon
hydrochloride
hydrogel
hydrogen
hydromorphone
hydroxide
hydroxyethyl
hydroxyl
ice
ingredient
ink
insect
insecticide
insulin
interbody
iodidea
ion
iron
item
ivory
juice
keratin
keratinocyte
kernel
kerosene
ketone
lactose
lamina
latex
leaf
leather
lime
limestone
linoleum
lipid
liquid
lubricant
luggage
lye
lyophilizate
lysine
magnesium
mannitol
manure
marble
martensite
meat
medicament
medicine
membrane
mesalamine
metal
methane
methanol
methyl
methylbenzyl
methylene
microorganism
milk
mineral
mixture
molecule
mollusk
mud
myocardium
natural
nickel
niobium
nitro
nitrogen
nitrous
nucleotide
nut
nylon
object
oil
oligonucleotide
ore
organ
organic
organism
oxide
oxygen
paclitaxel
paint
paper
particle
particulate
pathogen
peanut
pearl
peel
peptide
pet
phosphate
phosphite
phospholipid
photograph
piperazine
pith
planet
plant
plasma
plaster
plastic
platinum
plaything
plenum
plywood
pollen
polyamino
polycarbonate
polyester
polyethylene
polyglycolide
polymer
polynucleotide
polypeptide
polyphosphate
polyphosphite
polystyrene
polyurethane

porcelain
potassium
poultry
powder
primer
prodrug
progeny
propane
propellant
propylene
protein
pulp
pvc
quartz
quinacrine
rebaudioside
reed
refrigerant
render
resin
reticulocyte
rock
root
rubber
saline
salt
sand
sap
sash
sawdust
seawater
seed
shampoo
sheet
silica
silicon
silicone
silver
slag
slate
sludge
slurry
slush
smoke
soap
sodium
soil
solid
solute
solvent
soy
soybean
specimen
spider
sponge
starch
steam
steel
stem
steroid
stockpile
stone
stream
substance
substrate
sucrose
sugar
sulfate
sulfide
sulfur
sunlight
surfactant
sweetener
syrup
tantalum
tarpaulin
tea
termite
tetracycline
textile
thatch
thermoplastic
thread
tile
timber
time
tin
tissue
titanium
tobacco
transferase
tree
triethylamine
trifluoromethyl
triglyceride
tuft
tungsten
uranium
urethane
urine
vaccine
vantablack
vapor
vegetable
veneer
vinyl
waffle
wallpaper
warp
waste
water
wax
weed
wheat
whiteout
wire
wood
wool
yarn
yeast
zinc

References

[1] Ullman, D. G., 2003, *The Mechanical Design Process*, 3rd ed., McGraw-Hill, Boston, MA.

[2] Fu, K., 2012, "Discovering and Exploring Structure in Design Databases and Its Role in Stimulating Design by Analogy," Ph.D. thesis, Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, PA.

[3] Murphy, J. T., 2011, "Patent-Based Analogy Search Tool for Innovative Concept Generation," Ph.D. thesis, The University of Texas at Austin, Austin, TX.

[4] World Intellectual Property Organization, 2017, "World Intellectual Property Indicators 2017," World Intellectual Property Organization, Geneva, Switzerland.

[5] Casakin, H., and Goldschmidt, G., 1999, "Expertise and the Use of Visual Analogy: Implications for Design Education," Des. Stud., **20**(2), pp. 153–175.

[6] Markman, A. B., Wood, K. L., Linsey, J. S., Murphy, J. T., and Laux, J. P., 2009, *Supporting Innovation by Promoting Analogical Reasoning*, Oxford University Press, New York.

[7] Tseng, I., Moss, J., Cagan, J., and Kotovsky, K., 2008, "The Role of Timing and Analogical Similarity in the Stimulation of Idean Generation in Design," Des. Stud., **29**(3), pp. 203–221.

[8] Song, H., Lopez, R., Fu, K., and Linsey, J., 2017, "Characterizing the Effects of Multiple Analogs and Extraneous Information for Novice Designers in Design-by-Analogy," ASME J. Mech. Des., **140**(3), p. 031101.

[9] Linsey, J. S., Wood, K. L., and Markman, A. B., 2008, "Modality and Representation in Analogy," Artif. Intell. Eng. Des. Anal. Manuf., **22**(02), pp. 85–100.

[10] Linsey, J. S., Markman, A. B., and Wood, K. L., 2012, "Design by Analogy: A Study of the WordTree Method for Problem Re-Representation," ASME J. Mech. Des., **134**(4), p. 041009.

[11] Cheong, H., and Shu, L. H., 2014, "Retrieving Causally Related Functions From Natural-Language Text for Biomimetic Design," ASME J. Mech. Des., **136**(8), p. 081008.

[12] Lucero, B., Turner, C. J., and Linsey, J., 2016, "Design Repository & Analogy Computation Via Unit Language Analysis (DRACULA) Repository Development," ASME Paper No. DETC2015-46640.

[13] Sanaei, R., Lu, W., Blessing, L. T. M., Otto, K. N., and Wood, K. L., 2017, "Analogy Retrieval Through Textual Inference," ASME Paper No. DETC2017-67943.

[14] Montecchi, T., Russo, D., and Liu, Y., 2013, "Searching in Cooperative Patent Classification: Comparison Between Keyword and Concept-Based Search," Adv. Eng. Inf., **27**(3), pp. 335–345.

[15] Mukherjea, S., Bamba, B., and Kankar, P., 2005, "Information Retrieval and Knowledge Discovery Utilizing a Biomedical Patent Semantic Web," IEEE Trans. Knowl. Data Eng., **17**(8), pp. 1099–1110.

[16] Ahmed, S., Wallace, K. M., and Blessing, L. T. M., 2003, "Understanding the Differences Between How Novice and Experienced Designers Approach Design Tasks," Res. Eng. Des.—Theory Appl. Concurrent Eng., **14**(1), pp. 1–11.

[17] Fricke, G., 1996, "Successful Individual Approaches in Engineering Design," Res. Eng. Des.-Theory Appl. Concurrent Eng., **8**(3), pp. 151–165.

[18] Bjorklund, T. A., 2013, "Initial Mental Representations of Design Problems: Differences Between Experts and Novices," Des. Stud., **34**, pp. 135–160.

[19] Cross, N., 2004, "Expertise in Design: An Overview," Des. Stud., **25**(5), pp. 427–441.

[20] Razzouk, R., and Shute, V., 2012, "What is Design Thinking and Why is it Important?," Rev. Educ. Res., **82**(3), pp. 330–483.

[21] Adams, R. S., Turns, J., and Atman, C. J., 2003, "What Could Design Learning Look Like?," Design Thinking Research Symposium (DTRS), Sydney, Austrailia, Nov. 17–19.

[22] Akin, Ö., 1990, "Necessary Conditions for Design Expertise and Creativity," Des. Stud., **11**(2), pp. 107–113.

[23] Atilola, O., Tomko, M., and Linsey, J. S., 2016, "The Effects of Representation on Idea Generation and Design Fixation: A Study Comparing Sketches and Function Trees," Des. Stud., **42**, pp. 110–136.

[24] Kang, I. S., Na, S. H., Kim, J., and Lee, J. H., 2007, "Cluster-Based Patent Retrieval," Inf. Process. Manage., **43**, pp. 1173–1182.

[25] U.S. Patent and Trademark Office Electronic Information Products Division—Patent Technology Monitoring Team (PTMT), 2018, "U.S. Patent Activity Calendar Years 1790 to the Present," U.S. Patent and Trademark Office, Alexandria, VA, accessed Mar. 12, 2018, https://www.uspto.gov/web/offices/ac/ido/oeip/taf/h_counts.htm

[26] Song, B. Y., and Luo, J. X., 2017, "Mining Patent Precedents for Data-Driven Design: The Case of Spherical Rolling Robots," ASME J. Mech. Des., **139**(11), p. 111420.

[27] Kemp, C., and Tenenbaum, J. B., 2008, "The Discovery of Structural Form," Proc. Natl. Acad. Sci. U. S. A., **105**(31), pp. 10687–10692.

[28] Paatero, P., and Tapper, U., 1994, "Positive Matrix Factorization—A Nonnegative Factor Model With Optimal Utilization of Error-Estimates of Data Values," Environmetrics, **5**(2), pp. 111–126.
[29] Lee, D. D., and Seung, H. S., 1999, "Learning the Parts of Objects by Non-Negative Matrix Factorization," Nature, **401**(6755), pp. 788–791.
[30] Pauca, V. P., Shahnaz, F., Berry, M. W., and Plemmons, R. J., 2004, "Text Mining Using Non-Negative Matrix Factorizations," Fourth SIAM International Conference on Data Mining, Lake Buena Vista, FL, Apr. 22–24, pp. 452–456.
[31] Xu, W., Liu, X., and Gong, Y., 2003, "Document Clustering Based on Non-Negative Matrix Factorization," 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval (SIGIR '03), Toronto, ON, Canada, July 28–Aug. 1, pp. 267–273.
[32] Choo, J., Lee, C., Reddy, C. K., and Park, H., 2013, "UTOPIAN: User-Driven Topic Modeling Based on Interactive Nonnegative Matrix Factorization," IEEE Trans. Vis. Comput. Graph, **19**, pp. 1992–2001.
[33] Dulebohn, J. H., and Hoch, J. E., 2017, "Virtual Teams in Organizations," Human Res. Manag. Rev., **27**(4), pp. 569–574.
[34] Kim, J., and Park, H., 2008, "Toward Faster Nonnegative Matrix Factorization: A New Algorithm and Comparisons," Eighth IEEE International Conference on Data Mining (ICDM), Pisa, Italy, Dec. 15–19, pp. 353–362.
[35] Kim, J., and Park, H., 2011, "Fast Nonnegative Matrix Factorization: An Active-Set-Like Method and Comparisons," SIAM J. Sci. Comput., **33**(6), pp. 3261–3281.
[36] Cichocki, A., and Phan, A. H., 2009, "Fast Local Algorithms for Large Scale Nonnegative Matrix and Tensor Factorizations," IEICE Trans. Fundamentals Electron. Commun. Comput. Sci., **E92a**(3), pp. 708–721.
[37] Lim, W., Du, R., and Park, H., 2018, "CoDiNMF: Co-Clustering of Directed Graphs Via NMF," Conference on Artificial Intelligence (AAAI18), New Orleans, LA, Feb. 2–7, pp. 3611–3618.
[38] Fu, K., Cagan, J., Kotovsky, K., and Wood, K., 2013, "Discovering Structure in Design Databases Through Functional and Surface Based Mapping," ASME J. Mech. Des., **135**(3), p. 031006.
[39] Greene, D., O'Callaghan, D., and Cunningham, P., 2014, "How Many Topics? Stability Analysis for Topic Models," *Joint European Conference on Machine Learning and Knowledge Discovery in Databases, ECML PKDD 2014*, Springer, Berlin, pp. 498–513.
[40] Kuang, D., and Park, H., 2013, "Fast Rank-2 Nonnegative Matrix Factorization for Hierarchical Document Clustering," 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL, Aug. 11–14, pp. 739–747.