Proceedings of the ASME 2018 International Design Engineering
Technical Conferences and Computers and Information in Engineering Conference
IDETC/CIE 2018
August 26-29, 2018, Quebec City, Quebec, Canada

DETC2018-85591

APPROACHES FOR SUPPORTING EXPLORATION FOR ANALOGICAL INSPIRATION WITH BEHAVIOR, MATERIAL AND COMPONENT BASED STRUCTURAL REPRESENTATIONS OF PATENT DATABASES

Hyeonik Song
School of Mechanical Engineering
Georgia Institute of Technology
Atlanta, GA, USA

Katherine Fu, PhD
School of Mechanical Engineering
Georgia Institute of Technology
Atlanta, GA, USA

ABSTRACT

This paper presents an explorative-based computational methodology to aid the analogical retrieval process in design-by-analogy practice. The computational methodology, driven by Non-negative Matrix Factorization (NMF), iteratively builds a hierarchical repositories of design solutions within which clusters of design analogies can be explored by designers. In the work, the methodology has been applied on a large repository of mechanical design related patents, processed to contain only component-, behavior-, or material-based content, to demonstrate that unique and valuable attribute-based analogical inspiration can be discovered from different representations of patent data. For explorative purposes, the hierarchical repositories have been visualized with a three-dimensional hierarchical structure and two-dimensional bar graph structure, which can be used interchangeably for retrieving analogies. This paper demonstrates that the explorative-based computational methodology provides designers an enhanced control over design repositories, empowering them to retrieve analogical inspiration for design-by-analogy practice.

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 support to retrieve analogies from an electronic patent database [2, 3]. The patent database is deemed an ideal design repository for its innovative ideas of various application fields 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 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.

1.1 Prior Studies in Design-by-Analogy

Design-by-analogy has been an active research area [2, 5-8], among which studies have focused on understanding the effects of analogies on ideation and design outcomes. For instance, Linsey et al. explored how various types of representation of information affect 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 query-based approach or straightforward input-output method to retrieve analogies. In a case of WordTree method, designers input a word characterizing a design's functionality and are returned only 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 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

Expert designers exhibit several cognitive characteristics and abilities during design process. First, they have better spatial memory than novices, being able to process more information [16, 17]. In fact, they gather more information than novices to ideate design solutions for a given problem [18]. Second, not only do they gather information and ideate more, they can cognitively organize the information and conceptualize abstract ideas by viewing the problems objectively [19, 20]. Third, they 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 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 various 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 for the analogical exploration are introduced in Section 2.1.

1.3 Patent Database as Design Repository

The United States Patent database has several features that make it a suitable design repository for the design-by-analogy practice. The database contains prior design solutions that contain valuable knowledge and are deemed "patentable". Patentable ideas can be further defined as ideas that are "useful", meaning that the ideas are functional and operable, and "novel", meaning 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 promising 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 also 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 simultaneously presents a challenge for an effectively mining of patents. To address this challenge, various 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]. Fu et al. used Latent Semantic Analysis (LSA), a computational text analysis tool, to extract contextual similarities within patent documents and structure them based on surface and functional features [2]. Murphy used a Vector Space Model algorithm to evaluate functional analogies within patents [3]. Although 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.

1.4 Hierarchical Structure of Patent Data

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 (LDA) [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 the efficient NMF computation suggest its competency for the large corpus topic modeling [33-36].

Similar to most semantic techniques, it starts with transforming a corpus into a word-document matrix, in which each row represents a word from an entire corpus, each column represents a patent, and each matrix element represents the frequency of word occurring in the patent document. Mathematically, the word-document matrix is represented by $X \in \mathbb{R}_{+}^{m\times n}$, 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, X, into two nonnegative matrices, namely $W \in \mathbb{R}_{+}^{m\times k}$ and $H \in \mathbb{R}_{+}^{k\times n}$ such that $X \cong WH$. Here, W is word-topic matrix whose i^{th} topic column is represented by the weighted distribution of words. Similarly, H is topic-document matrix whose j^{th} document column is represented by the weighted distribution of the respective topics. The

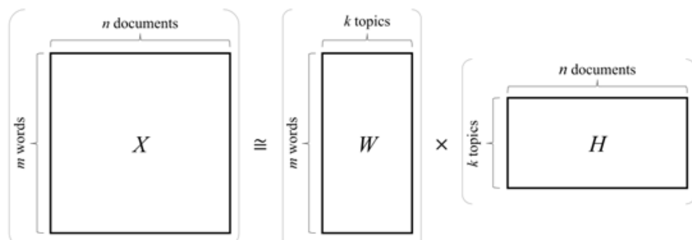


FIGURE 1. ILLUSTRATION OF NON-NEGATIVE MATRIX FACTORIZATION

matrix decomposition is illustrated in Figure 1.

2. METHODOLOGY

The patent data was processed to generate the cluster structures shown in Appendices A.1-3. First, word-document matrices of patents were prepared. Second, cluster structures of the patents were generated using Non-negative Matrix Factorization (NMF). Last, visual representations of the structures were generated to explore the cluster space. All computations 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 work involves retrieving patent data from a data storage system of United States Patent & Trademark Office (USPTO). The database consists of a bulk U.S. registered patents, each assigned to at least one classification term called Cooperative Patent Classification (CPC). All registered U.S. patents are categorized into one of 9 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 1,060 patents. This overall sample size (>1000 patents) was chosen to capture and analyze various analogical structures in the patent space with a goal of addressing the research gap in implementing computational DbA tool in a larger-scale design repository. Prior work by the authors has been with sample sizes of 100 patents. In future work, the authors hope to continue to scale up the sample size by orders of magnitude. For each patent document, only the words in the abstract, claims, and description sections were used, as they are the most representative of the patent's design features.

In addition to the patent documents, a design problem statement was also added to the corpus for generating the word-document matrix [37]. 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:

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 empower the designer to explore in a more tailored and efficient manner than before. The manipulation of the patent data set was done by manually compiling a list of words that characterize each analogical property, and then reducing the original matrix to contain only the rows of the listed words. The word lists were generated by first identifying what words appear in the patent corpora and picking those that apply to or describe any 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?

After a test run, words that appear too frequently were removed to distinguish one patent from another. As a result, word lists of component (709 words), behavior (262 words), and material (377 words) were compiled. 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 considered to be '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 [37]. 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 weighing was performed on the word-document matrices to assign higher weight for words that appear less frequently and vice versa.

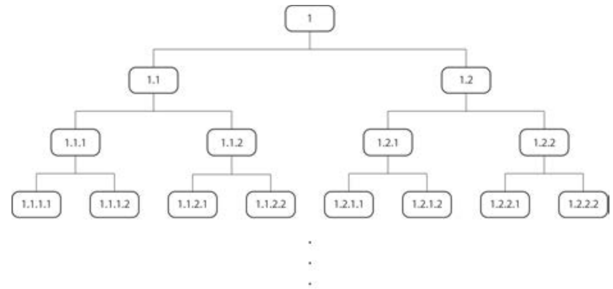


FIGURE 2. ILLUSTRATION OF RANK-2 NMF ITERATIONS

2.2 Processing with Topic Clustering Algorithm

The cluster structures of the three patent data sets were generated using NMF. As reviewed in Section 1.4, 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 [38] and thus questions the effectiveness of the topic clustering, especially for a large-scale data whose range of topics may vary exceedingly. To cope the challenge in this study, 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 Figure 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 [39], 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 10 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 [37]. 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 j^{th} column in the matrix contributes the most to the j^{th} 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 larger-scale cluster space. This section details two visualization methods used to analyze the clustered patent structures.

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

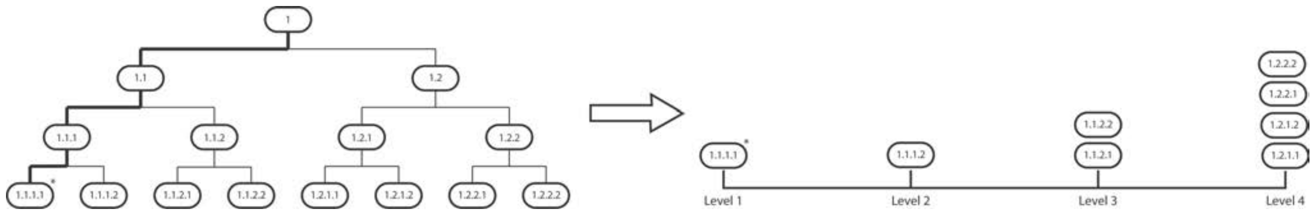


FIGURE 3. TRANSFORMATION OF 3D VISUALIZATION TO 2D VISUALIZATION

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 in 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 Appendices A1-3, was generated using MATLAB's 'digraph' function with 'force3' layout. '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 Figure 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 Figure 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 order of similarity from a starting leaf node (indicated with an * symbol) in Figure 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 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. 2D 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 AND DISCUSSION

The screenshots of the cluster structures are shown in Appendices A.1-3. 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 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 sub-functions 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 - one aspect is the material that the design is made of and the other is the raw material that the design uses.

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 analogical relationships among the clustered patents were evaluated using the computer-generated cluster labels. Figure 4 shows the simplified 2D visualization to analyze where three randomly chosen patents are assigned in the different cluster structures. For instance, 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. Clearly, each cluster represents a unique analogical relationship, suggesting that it could be beneficial for a designer who is seeking potential analogies for a desired design feature.

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.

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 spray 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, air, steam, carbon", suggesting that they use water. Although the two cluster labels represent different analogical relationships, they all have a common functionality of delivering fluids. This finding demonstrates that patents of common functionality are discovered by a common material used, which is not deemed a functional characteristic of the patented artifact such as the functional component or functional behavior in the previous examples. This finding is exciting, as a designer can discover patents of similar design attributes even with different analogical perspectives implying that the designer could retrieve a wide variety of analogical sources to gain design inspiration.

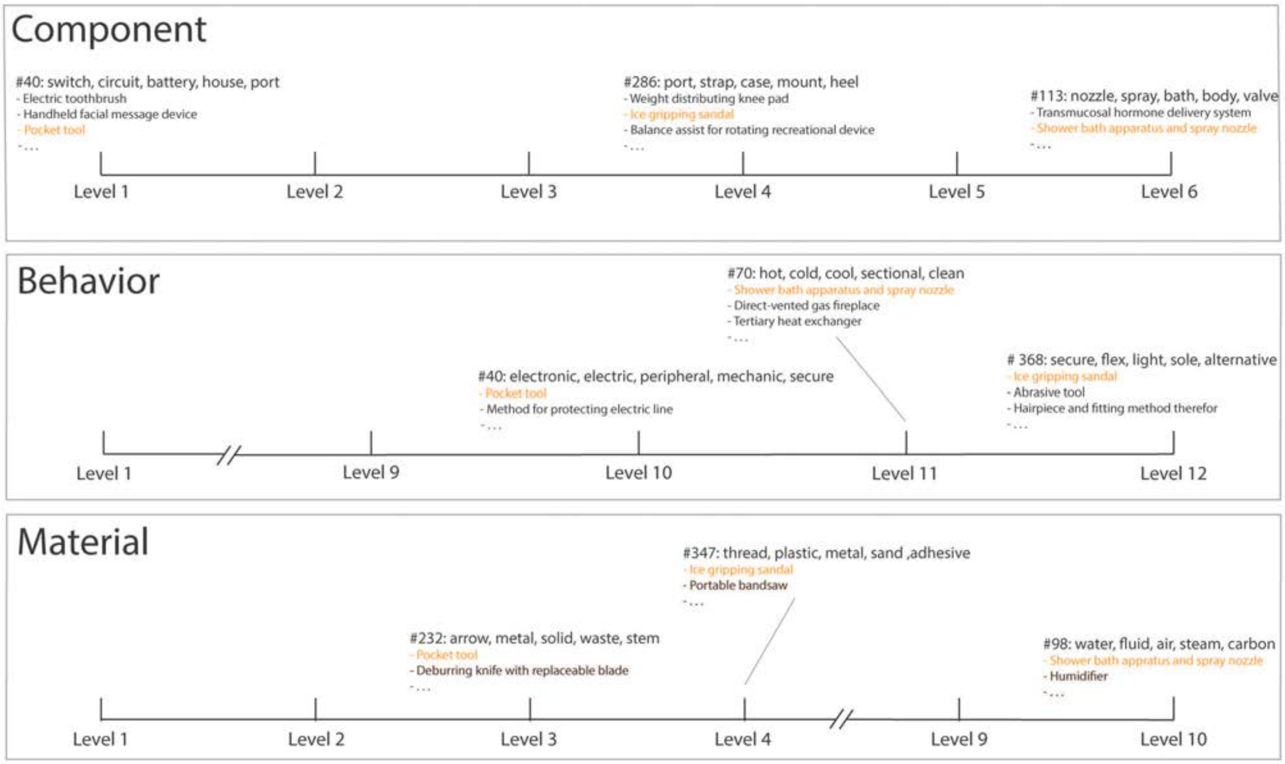


FIGURE 4. COMPARISON OF COMPONENT, BEHAVIOR, AND MATERIAL RESULTS

The analogical relationships discussed earlier provide a basis for user-tailored retrieval of analogies from patent data. Given the human motion energy collection design problem provided earlier, a designer could specifically explore for patents with energy converting components, patents with electrical behavior, or patents that are made of affordable natural materials to formulate several concepts depending on his/her design objective. Alternatively, the designer could freely browse adjacent or even distant patent clusters to intake various sources of analogical inspiration to conceptualize ideas from unexpected domains. The hierarchical structure of the patent data could be promising for designers as they can spontaneously make decisions and tailor their searches for the effective retrieval of analogies.

4. CONCLUSIONS

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 Non-negative Matrix Factorization (NMF) to generate a hierarchical cluster structure of the U.S. patent database where designers can visualize and evaluate analogies to solve given design problems. In the work, the user-controlled exploration technique has been implemented on the patent database, processed to contain only component-, behavior-, and material-based content, to demonstrate that different patent representations result in various analogical inspirations that are unique and valuable for ideation process. The interactive and multifaceted exploration tool empowers designers to effectively retrieve analogies from the design repository, facilitating the design-by-analogy practice.

ACKNOWLEDGMENTS

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.

REFERENCES

[1] D. G. Ullman, *The mechanical design process*, 3rd ed. Boston, Mass.: McGraw-Hill, 2003.

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

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

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

[5] H. Casakin and G. Goldschmidt, "Expertise and the use of visual analogy: implications for design education," *Design Studies,* vol. 20, pp. 153-175, 1999/03/01/ 1999.

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

[7] I. Tseng, J. Moss, J. Cagan, and K. Kotovsky, "The role of timing and analogical similarity in the stimulation of idean generation in design.," *Design Studies,* vol. 29, pp. 203-221, 2008.

[8] H. Song, R. Lopez, K. Fu, and J. Linsey, "Characterizing the Effects of Multiple Analogs and Extraneous Information for Novice Designers in Design-by-Analogy," *Journal of Mechanical Design,* vol. 140, 2017.

[9] J. S. Linsey, K. L. Wood, and A. B. Markman, "Modality and representation in analogy," *Ai Edam-Artificial Intelligence for Engineering Design Analysis and Manufacturing,* vol. 22, pp. 85-100, Spr 2008.

[10] J. S. Linsey, A. B. Markman, and K. L. Wood, "Design by Analogy: A Study of the WordTree Method for Problem Re-Representation," *Journal of Mechanical Design,* vol. 134, Apr 2012.

[11] H. Cheong and L. H. Shu, "Retrieving Causally Related Functions From Natural-Language Text for Biomimetic Design," *Journal of Mechanical Design,* vol. 136, Aug 2014.

[12] B. Lucero, C. J. Turner, and J. Linsey, "Design Repository & Analogy Computation via Unit Language Analysis (DRACULA) Repository Development," *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, 2015, Vol 1a,* 2016.

[13] R. Sanaei, W. Lu, L. T. M. Blessing, K. N. Otto, and K. L. Wood, "Analogy Retrieval through Textual Inference," *Proceedings of the Asme International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, 2017, Vol 2a,* 2017.

[14] T. Montecchi, D. Russo, and Y. Liu, "Searching in Cooperative Patent Classification: Comparison between keyword and concept-based search," *Advanced Engineering Informatics,* vol. 27, pp. 335-345, Aug 2013.

[15] S. Mukherjea, B. Bamba, and P. Kankar, "Information retrieval and knowledge discovery utilizing a biomedical patent Semantic Web," *Ieee Transactions on Knowledge and Data Engineering,* vol. 17, pp. 1099-1110, Aug 2005.

[16] S. Ahmed, K. M. Wallace, and L. T. M. Blessing, "Understanding the differences between how novice and experienced designers approach design tasks," *Research in Engineering Design-Theory Applications and Concurrent Engineering,* vol. 14, pp. 1-11, Feb 2003.

[17] G. Fricke, "Successful individual approaches in engineering design," *Research in Engineering Design-Theory Applications and Concurrent Engineering,* vol. 8, pp. 151-165, 1996.

[18] T. A. Bjorklund, "Initial mental representations of design problems: Differences between experts and novices," *Design Studies,* vol. 34, pp. 135-160, Mar 2013.

[19] N. Cross, "Expertise in design: an overview," *Design Studies,* vol. 25, pp. 427-441, Sep 2004.

[20] R. Razzouk and V. Shute, "What is design thinking and why is it important? (vol 82, pg 330, 2012)," *Review of Educational Research,* vol. 82, pp. 483-483, Dec 2012.

[21] J. T. Robin S. Adams, Cynthia J. Atman, "What could design learning look like?," in *Design Thinking Research Symposium*, Sydney, Austrailia, 2003.

[22] Ö. Akin, "Necessary conditions for design expertise and creativity," *Design Studies,* vol. 11, pp. 107-113, 1990/04/01/ 1990.

[23] O. Atilola, M. Tomko, and J. S. Linsey, "The effects of representation on idea generation and design fixation: A study comparing sketches and function trees," *Design Studies,* vol. 42, pp. 110-136, Jan 2016.

[24] I. S. Kang, S. H. Na, J. Kim, and J. H. Lee, "Cluster-based patent retrieval," *Information Processing & Management,* vol. 43, pp. 1173-1182, Sep 2007.

[25] (March 12). *U.S. Patent Activity Calendar Years 1790 to the Present*. Available: https://www.uspto.gov/web/offices/ac/ido/oeip/taf/h_counts.htm

[26] B. Y. Song and J. X. Luo, "Mining Patent Precedents for Data-Driven Design: The Case of Spherical Rolling Robots," *Journal of Mechanical Design,* vol. 139, Nov 2017.

[27] C. Kemp and J. B. Tenenbaum, "The discovery of structural form," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 105, pp. 10687-10692, Aug 5 2008.

[28] P. Paatero and U. Tapper, "Positive Matrix Factorization - a Nonnegative Factor Model with Optimal Utilization of Error-Estimates of Data Values," *Environmetrics,* vol. 5, pp. 111-126, Jun 1994.

[29] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature,* vol. 401, pp. 788-791, Oct 21 1999.

[30] V. P. Pauca, F. Shahnaz, M. W. Berry, and R. J. Plemmons, "Text mining using non-negative matrix factorizations," *Proceedings of the Fourth SIAM International Conference on Data Mining,* pp. 452-456, 2004.

[31] W. Xu, X. Liu, and Y. Gong, "Document clustering based on non-negative matrix factorization," presented at the Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, Toronto, Canada, 2003.

[32] J. Choo, C. Lee, C. K. Reddy, and H. Park, "UTOPIAN: user-driven topic modeling based on interactive nonnegative matrix factorization," *IEEE Trans Vis Comput Graph,* vol. 19, pp. 1992-2001, Dec 2013.

[33] R. D. Du, D. Kuang, B. Drake, and H. Park, "DC-NMF: nonnegative matrix factorization based on divide-and-conquer for fast clustering and topic modeling," *Journal of Global Optimization,* vol. 68, pp. 777-798, Aug 2017.

[34] J. Kim and H. Park, "Toward Faster Nonnegative Matrix Factorization: A New Algorithm and Comparisons," *Icdm 2008: Eighth Ieee International Conference on Data Mining, Proceedings,* pp. 353-362, 2008.

[35] J. Kim and H. Park, "Fast Nonnegative Matrix Factorization: An Active-Set-Like Method and Comparisons," *Siam Journal on Scientific Computing,* vol. 33, pp. 3261-3281, 2011.

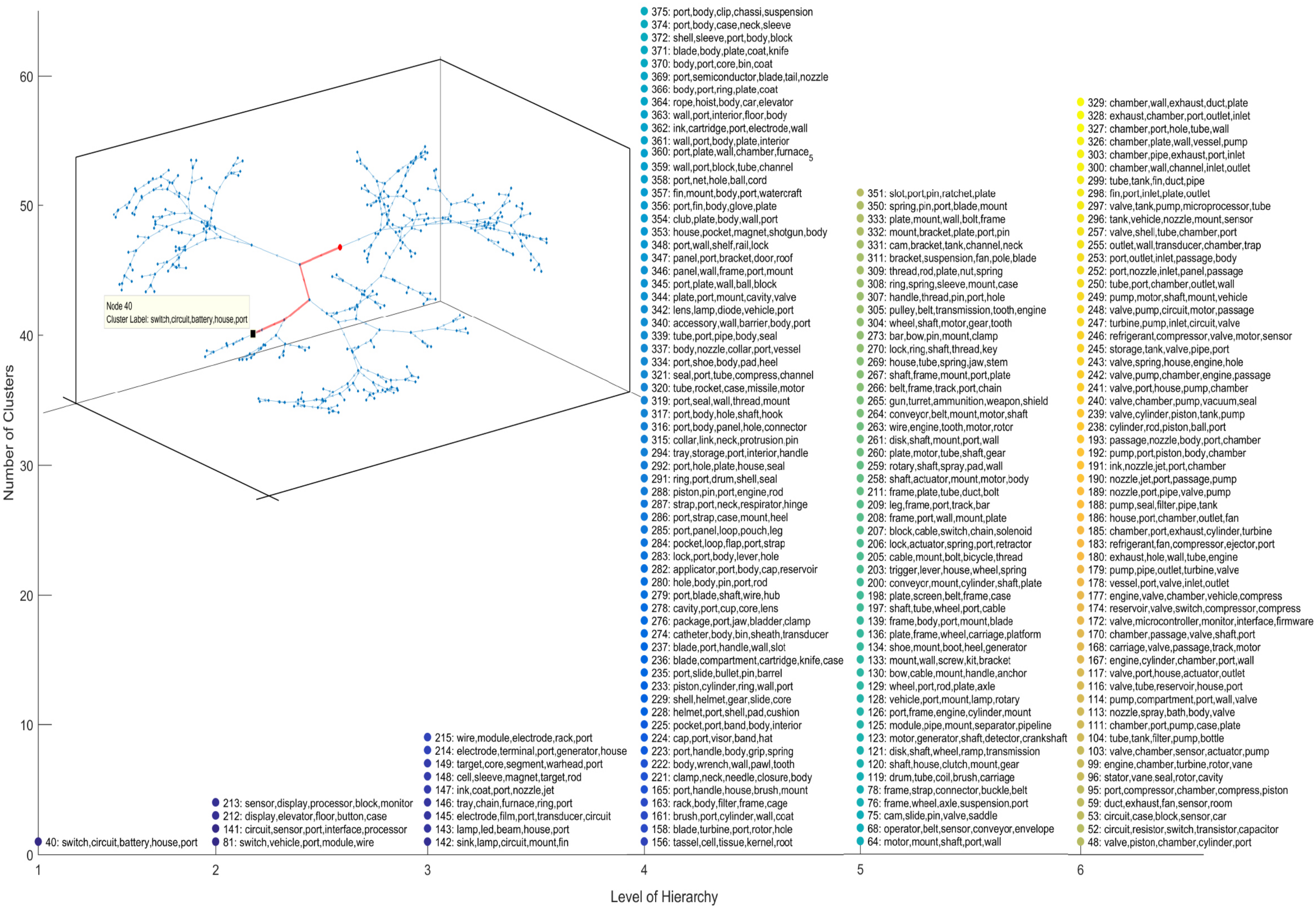
[36] A. Cichocki and A. H. Phan, "Fast Local Algorithms for Large Scale Nonnegative Matrix and Tensor Factorizations," *Ieice Transactions on Fundamentals of Electronics Communications and Computer Sciences,* vol. E92a, pp. 708-721, Mar 2009.

[37] K. Fu, J. Cagan, K. Kotovsky, and K. Wood, "Discovering Structure in Design Databases through Functional and Surface Based Mapping," *Proceedings of the Asme International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, 2011, Vol 9,* pp. 251-261, 2012.

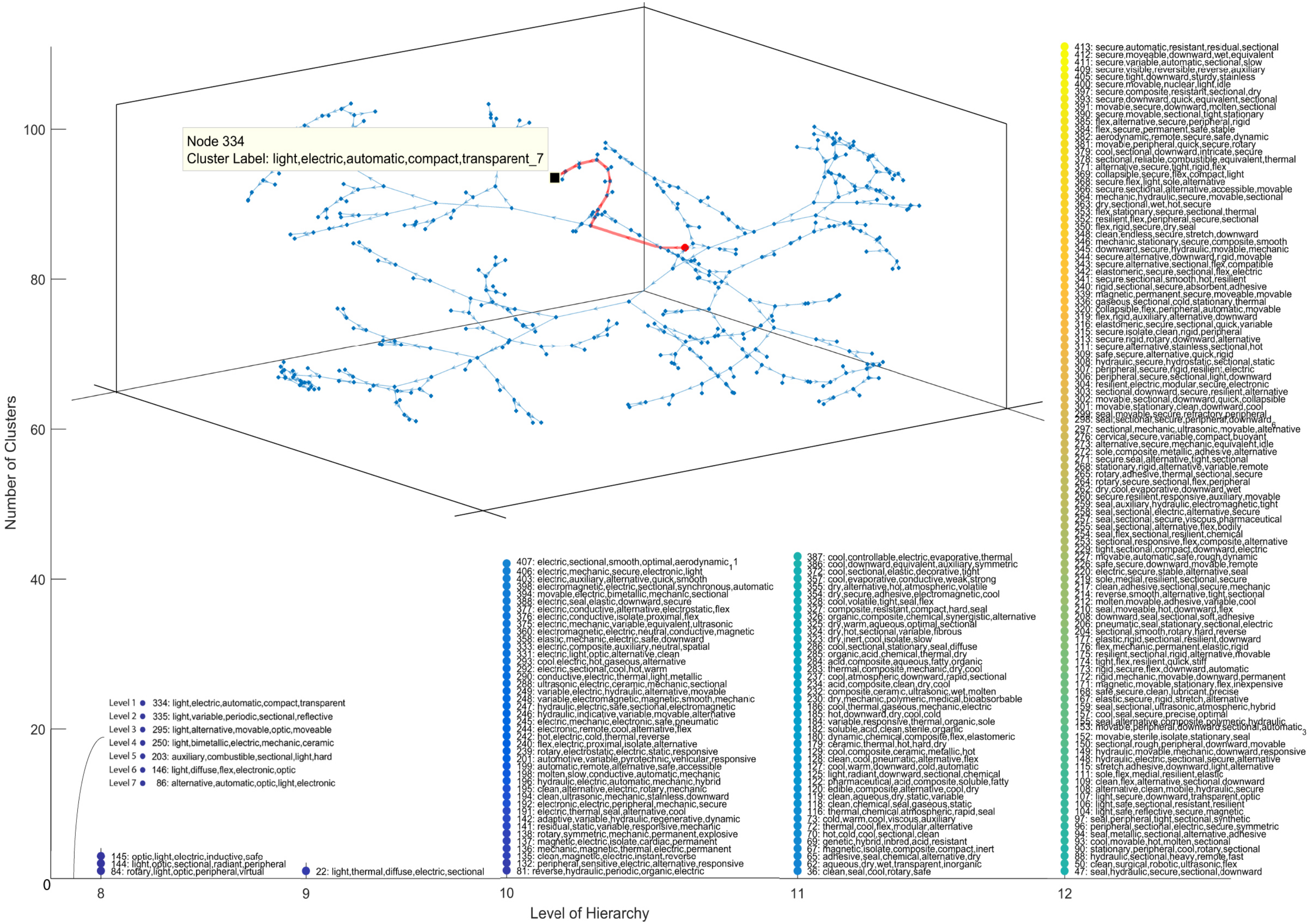
[38] D. Greene, D. O'Callaghan, and P. Cunningham, "How Many Topics? Stability Analysis for Topic Models," Berlin, Heidelberg, 2014, pp. 498-513.

[39] D. Kuang and H. Park, "Fast rank-2 nonnegative matrix factorization for hierarchical document clustering," presented at the Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, Chicago, Illinois, USA, 2013.

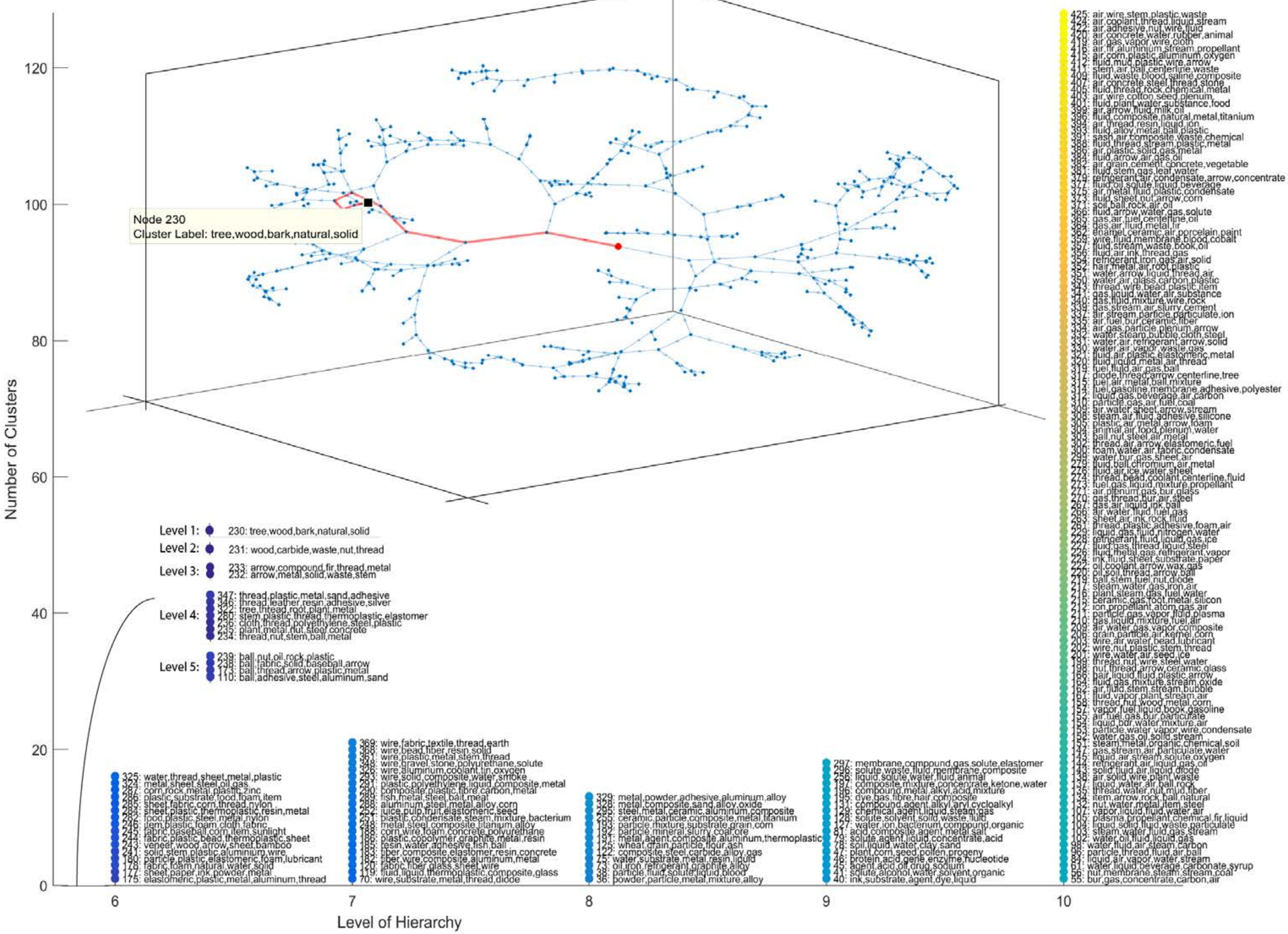
APPENDIX A.1: 3D and 2D Visualizations for Component Representation of Patent Database



APPENDIX A.2: 3D and 2D Visualizations for Behavior Representation of Patent Database



APPENDIX A.3: 3D and 2D Visualizations for Material Representation of Patent Database



APPENDIX B: List of 53 CPC Sub-Sections

Section	Subsection	Description
Section A: Human Necessities	1	Agriculture; forestry; animal husbandry; hunting; trapping; fishing
	41	Wearing apparel
	42	Headwear
	43	Footwear
	45	Hand or travelling articles
	46	Brushware
	47	Tables; desks; office furniture; cabinets; drawers; general details of furniture
	61	Medical or veterinary science; hygiene
	62	Life-saving; fire-fighting
	63	Sports; games; amusements
Section B: Performing Operation; Transporting	2	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	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; lining machines; typewriters; stamps
	60	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	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 subclasses of classes
	15	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
	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
	42	Ammunition; blasting