Discovering Structure in Design Databases Through Functional and Surface Based Mapping

Katherine Fu
Jonathan Cagan
Department of Mechanical Engineering,
Carnegie Mellon University,
Pittsburgh, PA 15213

Kenneth Kotovsky
Department of Psychology,
Carnegie Mellon University,
Pittsburgh, PA 15213

Kristin Wood
Engineering and Product Development Pillar,
Singapore University of Technology and Design,
Singapore, 138682 Republic of Singapore

This work presents a methodology for discovering structure in design repository databases, toward the ultimate goal of stimulating designers through design-by-analogy. Using a Bayesian model combined with latent semantic analysis (LSA) for discovering structural form in data, an exploration of inherent structural forms, based on the content and similarity of design data, is undertaken to gain useful insights into the nature of the design space. In this work, the approach is applied to uncover structure in the U.S. patent database. More specifically, the functional content and surface content of the patents are processed and mapped separately, yielding structures that have the potential to develop a better understanding of the functional and surface similarity of patents. Structures created with this methodology yield spaces of patents that are meaningfully arranged into labeled clusters, and labeled regions, based on their functional similarity or surface content similarity. Examples show that cross-domain associations and transfer of knowledge based on functional similarity can be extracted from the function based structures, and even from the surface content based structures as well. The comparison of different structural form types is shown to yield different insights into the arrangement of the space, the interrelationships between the patents, and the information within the patents that is attended to—enabling multiple representations of the same space to be easily accessible for design inspiration purposes. In addition, the placement of a design problem in the space effectively points to the most relevant cluster of patents in the space as an effective starting point of stimulation. These results provide a basis for automated discovery of cross-domain analogy, among other implications for creating a computational design stimulation tool. [DOI: 10.1115/1.4023484]

1 Introduction

There are many methodologies and philosophies for achieving effective engineering design, one proven approach to achieving innovative solutions being "design-by-analogy" [1]. Design-by-analogy is a process in which designers use design solutions from other domains in order to gain inspiration or insight for the design problem at hand. Design-by-analogy is becoming more popular with designers in industry. Some examples include applying MEMS (micro-electro-mechanical systems) technology to the manufacture of photovoltaic cells, using video game technology to inspire the control of a BMW, or using formula 1 vehicle suspension systems as inspiration for Nike shoe shock absorption technology [2–5]. This method remains challenging, however, due to the lack of a practical, efficient, procedural way to find these meaningful analogies. The work presented here attempts to make progress toward a solution to that challenge.

1.1 Understanding Design-by-Analogy Through Cognitive Studies. The way humans form concepts about design is important to understanding the use of analogies. Bloom found that when adult humans attempt to categorize artifacts in to artifact kinds, the function of the artifact is not the only important component in making this judgment but also the physical appearance. More importantly, he conjectures that the crucial factor in determining artifact kind lies in the intention of the creator of the artifact [6]. This work is directly related to humans' ability to find and use analogies in design. If designers find a design solution or technology with potential for analogical transfer (a source artifact) and categorize it as a particular artifact kind, their ability to transfer that knowledge to an alternate artifact kind or application may be hindered by previous categorization of the artifact.

Defeyter and German found that artifact concepts play an important part in problem solving, and that the intended purpose leads to their concept of artifact function [7]. Thus, if humans naturally struggle to repurpose artifacts when primed with their intended use, there may be implications for the conditions under which analogical transfer is more likely to occur.

Analogy and external stimuli in engineering design has been studied in a number of ways. Studies have been performed to understand how the introduction of analogies affects the ideation process and outcomes [8–11], with some studies specifically examining how the introduction of analogies with different levels of applicability to the design problem affects individual designers [12,13]. In addition, work has been done to better understand when and how the introduction of external stimuli to designers is most beneficial to design outcomes. For example, it has been shown that if subjects have "open goals" (i.e., unsolved problems) in mind when exposed to information that could be relevant to the design problem, those open goals can aid problem solving [14,15]; this open goal effect is achieved by giving subjects supplemental valuable information, or hints consisting of distant or unobvious information, only after solving has already begun.

Tseng et al. further studied the effect of open goals in combination with manipulating the type of external stimuli introduced. They found that giving subjects information that was analogous but distantly related to the design problem caused them to produce more solutions with more diversity and a higher level of novelty when open goals existed; in the absence of open goals (i.e., prior to the introduction of the problem to be solved), highly similar analogous information was more easily applied than distantly related analogous information [13].

Contributed by Design Theory and Methodology Committee of ASME for publication in the Journal of Mechanical Design. Manuscript received May 14, 2012; final manuscript received January 2, 2013; published online February 20, 2013. Assoc. Editor: Bernard Yannou.

Negative effects of introducing analogical information or examples are possible as well, a main one being design fixation [16–19], or the "blind adherence to a set of ideas or concepts limiting the output of conceptual design" [16]. Jansson and Smith showed that introducing examples can cause designers to generate solutions that mimic the examples, to the point of violating the design problem objectives [16]. Ward et al. showed that designers included aspects of examples in their solutions, even when explicitly told not to, implying that they have little control over the degree to which they are influenced by examples they see Refs. [20,21]. The extension of the work presented here is to facilitate the use of analogy to inspire designers, with the presupposition that these external stimuli would be introduced at the appropriate point in ideation, in the most helpful format and under the best conditions, as informed by the previous research outlined above.

1.2 Design Tools and Computational Design Aids. A significant area of research in engineering design is the creation of computational tools to aid designers during the design process. Stone and Wood created a functional basis in order to provide a universal language to facilitate functional modeling, a useful tool in the ideation process [22]. This functional basis work has been extended and adapted a great deal, one example of which is a biological functional basis [23]. This functional basis and language of design work is crucial to the work presented here, as it informs the exploration of functional interrelatedness of patents as compared to surface interrelatedness.

The U.S. patent database has been a focus of computational design aids for its convenient cataloguing of extant technology and engineering design. TRIZ (from Russian, the theory of inventive problem solving) uses heuristic rules (such as use of opposites) to help engineers overcome impasses in functional reasoning by searching through patents [24]. TRIZ and functional basis have been combined to create an axiomatic conceptual design model [25]. Souili et al. developed a method to identify candidates automatically using linguistic markers for use in TRIZ and inventive design method (IDM) [26,27]. Cascini and Russo created a method for automatically identifying the contradiction underlying a given technical system using textual analysis of patents for use in TRIZ [28]. Patent citation data have been used to find the interrelatedness between technologies, and the benefits of tapping into the technology knowledge base created by competitors within a particular design field [29]. Syntactic similarity between patent claims has been explored for the purpose of aiding in patent infringement research [30]. Patent repository tools and patent mining have been used to ascertain potential future market trends, recognize prolific inventors, and more, for business purposes. The mining of these patents included characterizing them by the number of citations, number of claims, average number of words per claim, number of classes that the patent spans, etc., [31]. Patent mapping methods of semantic content have been developed for the purpose of ascertaining patent portfolio overlap in mergers and acquisitions of companies [32]. Bohm et al. used the Design Repository at Oregon State University to perform a function based search using Chi matrix and morphological matrix techniques to find components that were present in concepts generated by hand, showing the potential for a computational design aid tool [33]. Our work focuses on using the textual content of the patents, which it is hoped will allow for richer outcomes. In addition, design repositories in general, not necessarily populated with patent data, have been explored as resources for designers, serving as ways to share and reuse designs to streamline the product design of complex engineering systems [34]. This previous design repository work involved storing CAD (computer aided design) models of components and assemblies for future design applications in a central database, allowing designers to save time and perhaps gain insight into previous models and designs. Our focus is on structuring design repositories and more open-ended analogical transfer.

Koch et al. created a tool called PatViz, which allows for visual exploration of iterative and complex patent searches and queries using all types of patent data, including full text. One graph view within this tool is created by the user in a guided process, not through an algorithm. There are three visualizations of interest within the tool called Patent Graph, which is a fully connected web of patents, and 3D image plane detector (IPD) Treemap, which is a 3D tree structure of the patents based on a predefined classification schema, and the Aggregation Tree, which is another tree view that deals with predefined adjustable hierarchies [35]. The important difference between the work of Koch et al. and the work presented here is that the structures within the PatViz tool are either predefined or user-defined classification schemes, while this work uses an exploration methodology to discover the best (and multiple different) structures to describe the set of patents. The form of the structure itself changes as the data being examined changes.

A BioMedical Patent Semantic Web was created by Mukherjea et al., which found semantic associations between important biological terms within biomedical patents and returned a ranked list of patent resources and a Semantic Web that displays the relationships between the important terms and between resources. This work was performed with the intent of aiding in avoiding patent infringement. The Semantic Webs are fully connected graphs with no imposed structure, and the data used only include the abstract of the patents being examined. In addition, the webs were not generated using a Bayesian inference approach [36]. Chakrabarti et al. used a topic model, which employs Bayesian inference to train a model on a small data set of documents and then automatically categorize the remaining documents into "topics," leading to a taxonomy or hierarchical structure [37]. That work does not explore structures other than hierarchies, and is not applied to the exploration of these structures as fodder for analogical design work.

As stated previously, it is logical to turn to the U.S. patent database as an effective repository of analogical or cross-domain design solutions. However, due to the size and complexity of the U.S. patent database, it is difficult to make it useful to designers. There have been many attempts to automate, aid, or streamline the search of the U.S. patent database. Theories like TRIZ and their resulting tools [24,25,38–45], and even the simple key word search on the United States Patent and Trademark Office (USPTO) website or Google Patents have attempted to make access to the information more streamlined, but it is still difficult to understand the characteristics relevant to a design problem within the "space" of patents [24]. Computational "innovation support tools" marketed to businesses and innovators have also been developed [46,47], as well as a number of other research driven design support tools and methodologies [48–52]. These methods and tools rely heavily on the users to generate the terms or analogies of their own accord and dredge through search results. Psychology literature indicates that retrieval of far-field analogies is cognitively difficult [53], and remindings tend to be limited by surface similarity [54], meaning the probability of retrieving surface dissimilar analogies is low. With a way to extract the interrelatedness and interconnectedness of patents in the space and in addition, with respect to a specific design problem, designers might be able to strategically choose which cross-domain designs to expose themselves to, or even traverse the space in a more intentional and meaningful exploratory way. By allowing for more efficient and insightful access to external analogical stimuli, designers have the potential to create more innovative design solutions. The algorithm and methodology behind discovering this interrelatedness, or structure, within a patent space, or design repositories in general, is presented next.

1.3 Bayes and Discovering Structural Form. Bayesian models have been used to describe human cognition for centuries [55]. Jaynes describes human plausible reasoning as a calculation of the degree of plausibility of a particular hypothesis being true based on previous experience and common sense, and given the facts at hand, corresponding directly to the components that must be considered when calculating the posterior probability of a hypothesis being true given a set of data using Bayes rule [56]. This link between the Bayesian algorithm and human cognition is crucial to the motivation behind this methodology. It is hoped that the structures of patents will be closer to human structuring of data or information, and therefore more easily understood and useful.

Although finding the structure in data is not an easy or new problem, it has the potential to yield valuable insights if successful. Linneaus' discovered that living organisms are best described by a tree structure, or Mendeleev found the periodic structure of the elements [57]. More elementary to understanding and discovering structures, however, is clustering and categorization of data. Categorization is a topic that has been studied both in human cognition and in modeling human cognition [58–62]. While methods such as Latent Dirichlet Analysis (LDA) [63] and others have been used to categorize documents based on the text within them, extracting taxonomies and semantic similarity [63–66], LSA was chosen because it has been used successfully by Dong et al., the authors, and others in our field to track common understanding and design representations. LSA will be discussed in more detail, as it is a main component of the methodology presented for discovering structural form in a patent space. However, first, we review Bayesian reasoning.

Bayes Rule is a result of elementary probability theory. Given two random variables A and B, which, respectively, can take on values a and b, the following relationship describing the joint probabilities and a and b are true

$$P(a, b) = P(a|b)P(b) \quad (1)$$

$$P(a, b) = P(b|a)P(a) \quad (2)$$

Setting Eqs. (1) and (2) equal, we can rearrange them to be Bayes rule

$$P(b|a) = \frac{P(a|b)P(b)}{P(a)} \quad (3)$$

There are four terms in Bayes Rule, each with an important significance. For example, consider a problem in which an agent is trying to infer a process responsible for generating data, d. h is a hypothesis about what this process could be [55]. Bayes Rule for this scenario is written as

$$P(h|d) = \frac{P(d|h)P(h)}{P(d)} \quad (4)$$

where the following can be defined:

(1) The "*prior probability*"—$P(h)$, the probability that agent ascribes h is the true generating process, independent of data d.
(2) The "*posterior probability*"—$P(h/d)$, how agent should change beliefs in light of d—or, the degree of belief in h conditioned on observation of d.
(3) The "*likelihood*"—$P(d/h)$, the probability of the data given the hypothesis, reweights each hypothesis by how well it predicts the data.
(4) The "*marginal probability*"—$P(d)$, the probability distribution associated with d, calculated by summing over the other variable(s) in the joint distribution, where

$$P(d) = \Sigma_{h' \epsilon H} P(d|h')P(h') \quad (5)$$

and h' is an alternative hypothesis, H is the set of all hypotheses considered. The posterior probability, the left hand side of Bayes Rule, is calculated using the other three terms described above [55]. Kemp and Tenenbaum use this formula to calculate the probability that the data has structure S and form F given data D. A form is defined by the graph grammar that is used to create it. These forms include a partition, chain, order, ring, tree, hierarchy, grid, and cylinder, described in more detail in Sec. 1.4. These structures originate from psychology literature [67] and appear in formal models in many different research efforts [60,68–79]. One example from the work of Inhelder and Piaget is the classification scheme that children use in simple logic operations—which is based on a tree structure and an order. Kemp and Tenenbaum argue that the structural forms included in the algorithm are often and commonly found, are "useful for describing the world and that they spring to mind naturally when scientists seek formal descriptions of a domain" [57].

A structure S is a particular instantiation of a form F. To be clear, a graph of data D with a certain form can be represented by a number of different configurations, or structures. The three terms that go into calculating this posterior probability, which serves as the score of a particular structural form within the algorithm, were chosen and calculated as follows [57]:

$$P(S, F|D) \propto P(D|S)\, P(S|F)\, P(F) \quad (6)$$

where

(1) $P(F)$, the prior on the space of forms, is a uniform distribution over the forms under consideration.
(2) $P(S|F)$, the prior on the structures, favors graphs where k, the number of clusters, is small: $P(S|F) \propto \theta^k$ if S is compatible with F, and $P(S/F) = 0$ otherwise; here, $\theta = e^{-3}$.
(3) $P(D|S)$, the likelihood, measures how well the structure S accounts for the data D. $P(D|S)$ will be high if the features in D vary smoothly over the graph S, that is, if entities nearby in S tend to have similar feature values.
(4) The normalizing constant, the marginal probability, is calculated using set theory, as a sum of the products of the number of F-structures with k occupied cluster nodes and the number of ways to partition n elements into k nonempty sets.

1.4 Structural Form Descriptions.

The possible structural forms to consider with the algorithm are described in Fig. 1 [57,79]. For each type of structural form, a form is shown and a language of generative rules (graph grammar) is given that describes how the structural form is generated.

1.5 Latent Semantic Analysis.

In our work, Kemp and Tenenbaum's algorithm is combined with preprocessing and postprocessing using LSA, a computational text analysis tool that extracts contextual similarity of documents and words [64–66]. LSA has four main steps as follows:

(1) A word-by-document matrix is created, in which the columns are the individual text passages (here, the patents), the rows are the words that appear in the documents, and the cells are populated by a tally of the number of times each word appears in each document.
(2) An "entropy weighting" step is performed, a two-part transformation on the word-by-document matrix that gives a more accurate weighting of the word-type occurrences based on their inferred importance in the passages. For example, if a word occurs very often across all documents, it will have a low weight in the space, like words "the" and "a," etc.
(3) Singular value decomposition (SVD) is performed on the transformed matrix, with an output of three matrices (U, S, and V). U and V are orthonormal matrices whose rows and columns correspond, respectively, to the words and documents in the LSA space. S is a diagonal matrix of singular values. Dimensionality reduction of the LSA space can be performed by altering the S matrix to only contain the top n values along the diagonal, which can eliminate noise and lead to better results in analyses with large corpora.

However, due to the small size of the corpora used in the example in this paper (100 patents, although the methodology can in principle handle a much larger number), full dimensionality was maintained.

(4) The cosine similarity between documents can then be calculated by multiplying S and the transpose of V and calculating the dot product between all pairs of resulting vectors. This yields what is essentially a matrix of document-to-document coherence values. These values range from -1 to 1, where -1 signifies a perfect negative correlation, 1 signifies a perfect positive correlation, and 0 signifies that there is no correlation. Thus, if two documents were exactly the same, a value of 1 would be output for that cosine similarity [64–66].

2 Methodology

There are three parts to producing the structures presented in the Sec. 3 of this paper. First, LSA is used to preprocess the patents, producing patent similarity data. Second, a Bayesian inference algorithm devised by Kemp and Tenenbaum is used to discover structural forms in the patent data, using the output from LSA as input. Third, LSA is used again to create labels that describe the clusters of patents that have been created in the structure, based on word to document similarity calculations.

2.1 Preprocessing With LSA. LSA is used in this work to generate "similarity" data for input into the structural form discovery algorithm. Given an initial set of patents, the abstract and description of the patents are first parsed from HTM (HyperText Markup Language) text. Using a part-of-speech (POS) tagger, the verbs, adverbs, adjectives and nouns are tagged separately for each patent, and repeat words are included. The set of patents is reduced to only verbs for one data set, the functional data set, and only nouns for the second data set, the surface data set. This concept is derived from the work of Stone and Wood, in which the authors discuss how the part of speech of a word indicates its roll in design descriptions [22]. Verbs tend to describe **functionality** because they correspond to what something *does* or *should do*. Nouns tend to describe components, applications, or elements of a design, and thus are chosen here to represent **surface** attributes of the patents. Further work could include an expansion on this concept, including non-infinitive verb forms in the function data set and adjectives in the surface data set. The set of 100 patents used are enumerated in Appendix A including their U.S. patent number, title, and index number with respect to the results in this paper. These patents were randomly chosen, though prefiltered to include only mechanical patents, as classified within the U.S. Patent classification system as "body treatment and care, heating and cooling, material handling and treatment, mechanical manufacturing, mechanical power, static, and related arts." The patents were chosen randomly in order to exhibit the extendibility of this method to any set of patents, and its robustness in the face of limited intervention in the choosing of the initial subset of patents. The content of the structures and insights drawn from them will change depending on the set of patents that is chosen as input. It is hoped that eventually, this method will be extended to much larger datasets, to characterize very large portions of the patent database.

LSA, as described in Sec. 1.5, is then run on both data sets of the set of patents. Two symmetric cosine similarity matrices are produced, describing the similarity between the patents based on the functional or surface text within the patents. These matrices serve as the input similarity data for the algorithm for discovering structural form. Note that these data are not dissimilar in form from the similarity data used in Kemp and Tenenbaum's work, for example, the face recognition data [57]. Kemp and Tenenbaum did not use LSA to generate the data, however. This LSA preprocessing step does not alter the functionality of the algorithm devised by Kemp and Tenenbaum, but rather serves as an input data generation step.

2.2 Discovery of Structural Form. These structures are found across a wide variety of cognitive and other domains, as described in Sec. 1.3. We argue that if these structures are expressive of repeated descriptions of the natural world, then they are likely to be very relevant to designer thinking as well, making them a promising tool for stimulating design. This theory is tested in further work performed by the authors with a cognitive engineering design study [80,81]. The structures have the potential to indicate underlying relationships among design repository data, as they have done with many other example data sets as presented in Kemp and Tenenbaum's work, making them apt to uncover information that could be novel and inspirational to designers.

The algorithm for discovering structural form as it is applied to the LSA output patent similarity data includes the following steps [57,79]:

(1) Preprocess the feature data D by shifting the mean of the matrix to zero. Calculate normalized covariance matrix for D, defined as $(1/\mathrm{m})DD^{\mathrm{T}}$, where m is the number of features, or nonredundant nontrivial words included in the entire set of patents. Shifting the mean of D to zero normalizes the feature matrix to allow the calculated covariance to be comparable to the "empirical covariance."

(2) Find the form F and the structure S of that form that best capture the relationships between these patents by maximizing the posterior probability—the probability that the data has structure S and form F given data D; i.e., search for the structure S and form F that jointly maximize the scoring function $P(S, F|D)$. For example, the patents might best fit into the structural form of a tree.

(3) To identify the structure and form that maximize the posterior, a separate greedy search is run for each candidate form:
   - All patents are assigned to a single cluster.
   - The algorithm splits a cluster at each iteration, using a graph grammar that builds the structure (such as a tree) after each split.
   - Attempt to improve the score using several proposals, including proposals that move an entity from one cluster to another and proposals that swap two clusters.
   - The search concludes once the score can no longer be improved.

All eight forms, the partition, chain, order, ring, tree, hierarchy, grid, and cylinder, all shown in Fig. 1, were used as candidate forms. The output of this step of the methodology is the best structure (instantiation) of each candidate form, and the associated posterior probability. Using the posterior probability values, the best structure can be identified. In Sec. 3, we focus on the best structures as determined by the algorithm, though later work will explore the meaning and value of alternative structural representations of the data that are not the "best" in terms of posterior probabilities, but may have other useful meanings.[1]

2.3 Cluster Labeling With LSA. The third step in the methodology employs LSA once again. The purpose of the postprocessing is to create an automated way to be able to analyze the meaning of the connections between patents in the structures that are output with a characterization of the connections between or clustering of patents. If one attempted to understand the raw output from the algorithm, it would require extensive cross-referencing between patent documents and the structure, juggling many pieces of information at once. This postprocessing allows for a "snapshot" of the meaning of the connections to be seen. Latent Semantic Analysis is used to find the words in the LSA space that had the highest cosine similarity value to each patent

[1]The authors make the assumption that the algorithm produces valid results from a computational standpoint, as confirmed by the synthetic data analyses performed by Kemp and Tenenbaum. Due to the fact that the algorithm itself was unchanged, this is a valid assumption.

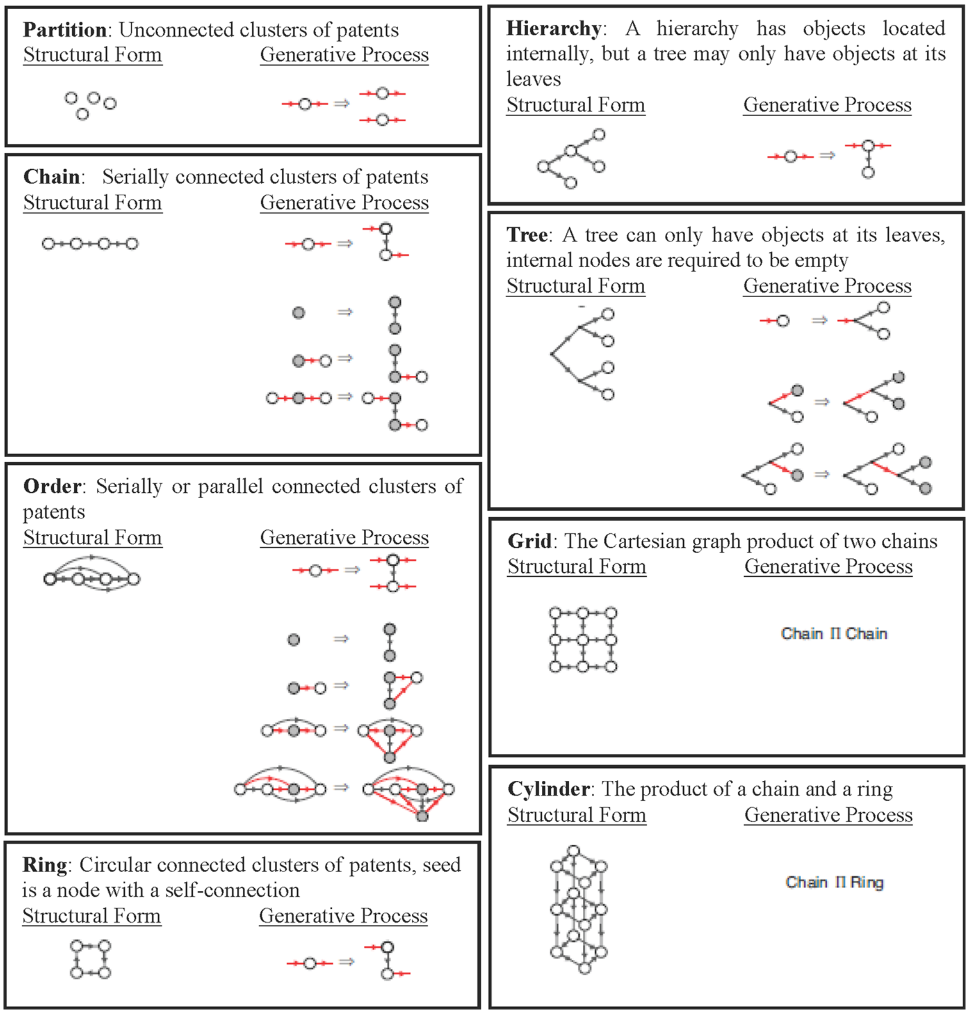

Fig. 1 Verbal and pictorial descriptions of structural forms and their generative processes

by multiplying the U, S, and V matrices. The output from this aforementioned calculation was then used in two different ways, described next. To summarize, the first method, the Highest Average Rank labeling method, generates labels that are based on the words that are the highest average rank to the set of patents in a cluster—often yielding a good, high level, more abstracted set of functionalities or terms that might be better suited for analogical mapping and transfer. However, the drawback of this method is that at times, the highest ranked terms in common to the patents in a cluster are too general, and do not supply specific enough information to be useful for analogical transfer. The second method, the Highest Cosine Similarity labeling method, generates labels that are based on the words that have the highest absolute cosine similarity value to each patent in the cluster. This yields labels that have very specific information to some or all patents, which can provide detail or context when the first method produces labels that are too broad or generic. However, this method can lead to labels that are not relevant to all patents in the cluster, which could be misleading in terms of the content and meaning of the cluster. The two methods are juxtaposed in the structures presented in Sec. 3, providing both sets of benefits, while mitigating the drawbacks of one another simultaneously.

2.3.1 Highest Average Rank Labeling. The first method of cluster labeling finds the words for the set of patents within a cluster that have the highest average ranks. Each patent has a column vector associated with it, comprised of the words in the LSA space ranked in order of descending cosine similarity value to that patent. The ranks for a word in the space are then tabulated across all patents in a particular cluster, and the average of these ranks is found. This is repeated for all words in the LSA space. The average ranks for the set of words in the cluster are then sorted in descending order, and the top words are used to label the cluster of patents within the graphs of the structural forms to allow them to be more easily interpretable.

2.3.2 Highest Cosine Similarity Labeling. The second method of cluster labeling finds the words for the set of patents with the highest cosine similarity values in the LSA space. The top twenty words and their corresponding cosine similarity values from each patent within a cluster are tabulated. The pooled top words from the patents within a cluster are then sorted by their cosine similarity values, and the top words are used to label the cluster of patents within the structural form graph. Again, the number of words used in the label is easily variable, but is chosen to be five initially in this work.

In the results presented next, the results from both labeling methods are displayed on the structures. The highest average rank method is noted as method 1, and the highest cosine similarity method is noted as method 2. There were some words that were excluded from the labels, as they were at the top of almost all lists of words, or deemed to be specific to patent description language rather than technology description language. These words included: use, provide, comprise, disclose, say, utilize, invention, invent, position, embodiment, describe, object, summary, background, material, fig, example, device, method, application, patent, assembly, end, extend, include, illustrate, refer, cause, mean.

2.4 Design Problem Placement. In addition to the 100 patents, a design problem was added into the LSA space as one of the documents. The design problem was one that has been used throughout this body of work by the authors, the text of which is the 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 kW h 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.

The purpose of this step was to locate what can be thought of as a "starting point" in the space, by calculating the average cosine similarity of the set of patents in each cluster to the design problem, and selecting the cluster with the highest average cosine similarity to the design problem as the starting point. This cluster is marked with a star in the figures shown in Sec. 3.

3 Results and Discussion

Figures 2–5 display the results of the methodology presented in Sec. 2, including the results of both labeling methods. Figures 2 and 3 show the best structure found by the algorithm, a hierarchy for both function based, meaning verb content from patents, and surface based, meaning noun content, patent data. Figures 4 and 5 show the third best structure found by the algorithm, a ring

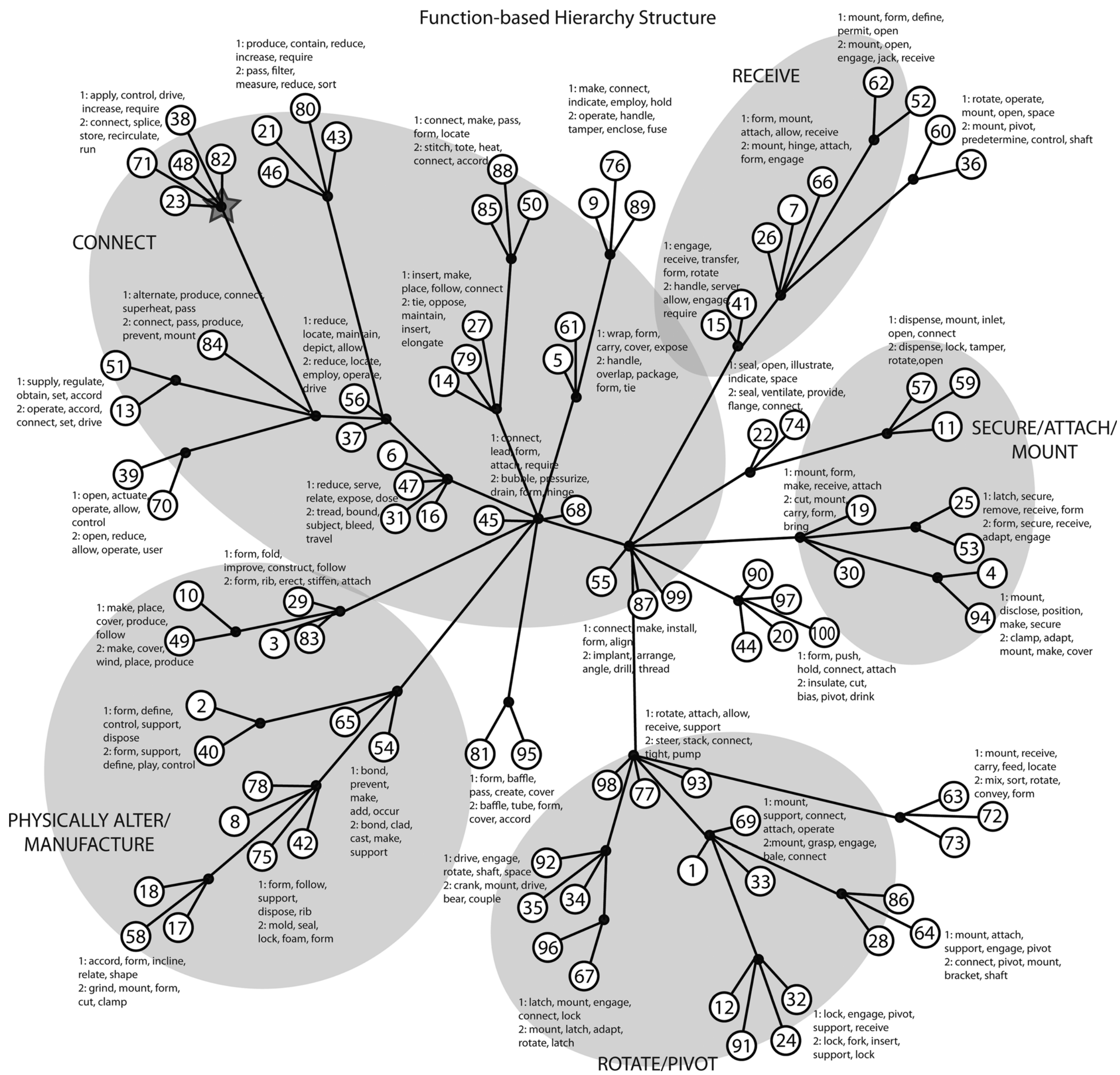

Fig. 2 Best function based structure, hierarchy showing regions of functionality

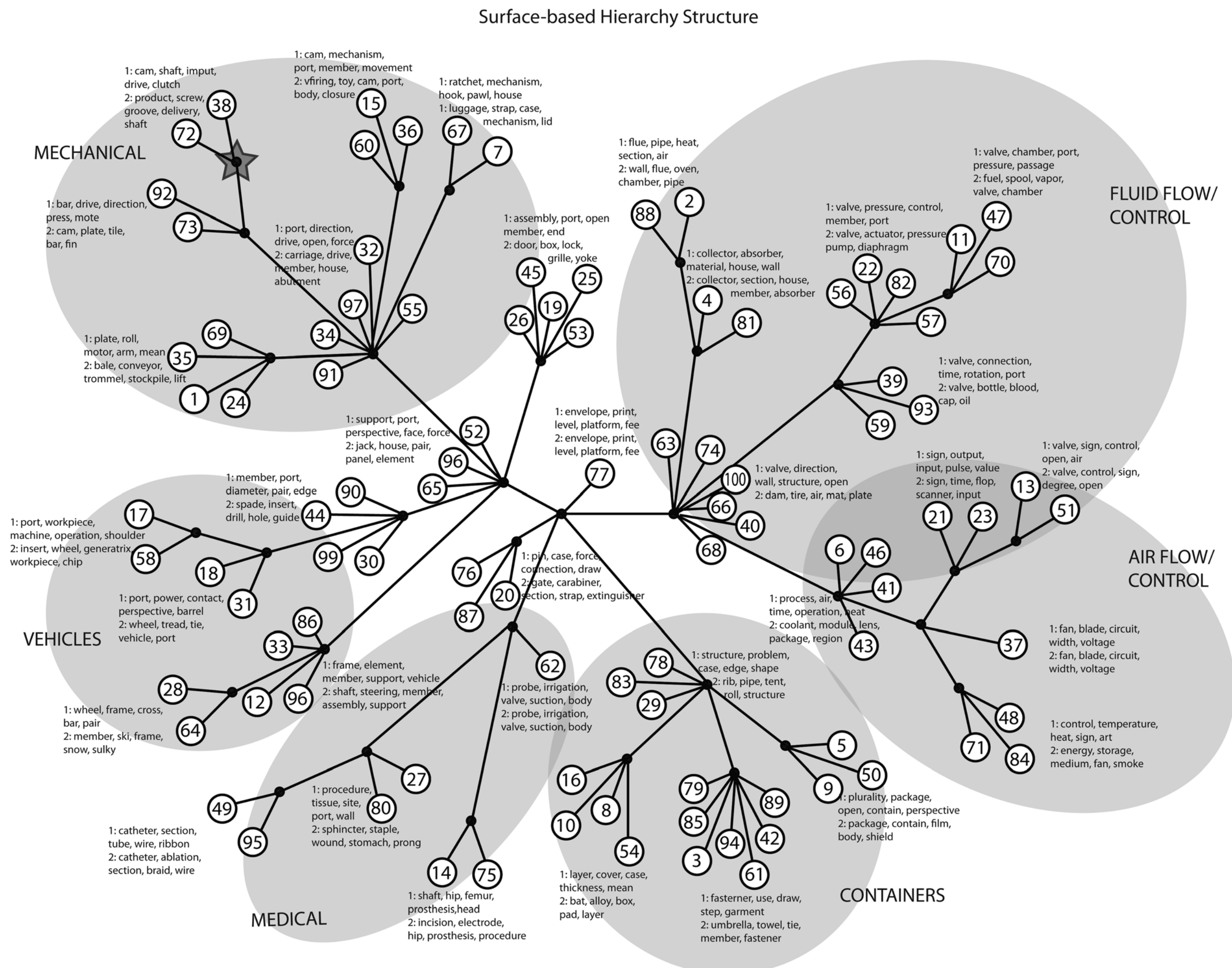

Fig. 3 Best surface based structure, hierarchy showing regions of surface content

structure for both function and surface based patent data. The third best structure was chosen to be included because the second best was a tree structure, which would not be as effective for demonstrating the diversity of structure types and their insights.

3.1 Regions. To start, the overall trends in the structures are discussed in order to become familiar with the results. Figures 2–5 display trends in the structures with respect to regions of functionality and surface content. These regions were extracted by the researcher, not computationally, which is a potential area of future research. The regions were extracted by examining the general trends of the labels from both methods, and determining areas of common function or surface content. In Fig. 2, it can be seen that there are five general regions of functionality, and these regions tend to align with the branches of the hierarchy structure. The regions include: receive, secure/mount/attach, rotate/pivot, physically alter/manufacture, and connect. While the patents in these clusters and regions do not in most cases have the main functionality of the region in which they are included, they do have subfunctionalities in common—which is to be expected when one considers the diversity of the immense patent database that will be represented in a random sampling of mechanical patents. These common subfunctionalities are exciting because they bring together patents from different application domains, potentially exposing a designer to technologies they would not have thought to be related, or even inspiring cross-domain analogical transfer of design information.

For example, the Fig. 2 shows patents 25, a horseshoe, and patent 19, Attachments in model airplanes, in the same region of "Secure/Attach/Mount" —in just looking at the titles of these patents, found in Appendix A, the reader can "at a glance" gauge that they have analogous functionality that might inspire a designer who is looking for different ways to affix something to something else. In a second example, Fig. 2 shows patent 62, an adjustable probe, and patent 41, automated apparatus and method for consolidating products for packaging, grouped together in the "Receive" region—where likely the probe was placed for *receiving* signals or information, and the packaging apparatus was placed for *receiving* goods to be packaged. This is an example of a grouping of patents into a region based on functionality that does not align with their area of application—which, again, could be beneficial to a design-by-analogy process seeking cross-industry transfer of knowledge. It is acknowledged that these regions, clusters, and labels are not perfect, and at times may be misrepresentative—an opportunity for improvement that will be addressed in future work.

Figure 3 displays the surface content regions extracted from the best surface based structure, a hierarchy. The regions that emerged were: fluid control/flow, air flow (really a subset of the previous), containers, medical, vehicles, and mechanical. Similar to the function based structure, the regions tend to align with the branches of the hierarchical structure, indicating that the structuring algorithm is extracting meaningful relationships between the patents and clusters of patents. The difference between the surface based structure regions and the function based structure regions is that, naturally, areas of surface similarity emerge, instead of functional similarity. For example, patent 27, wound site management and wound closure device, is grouped with patent 49, high performance braided

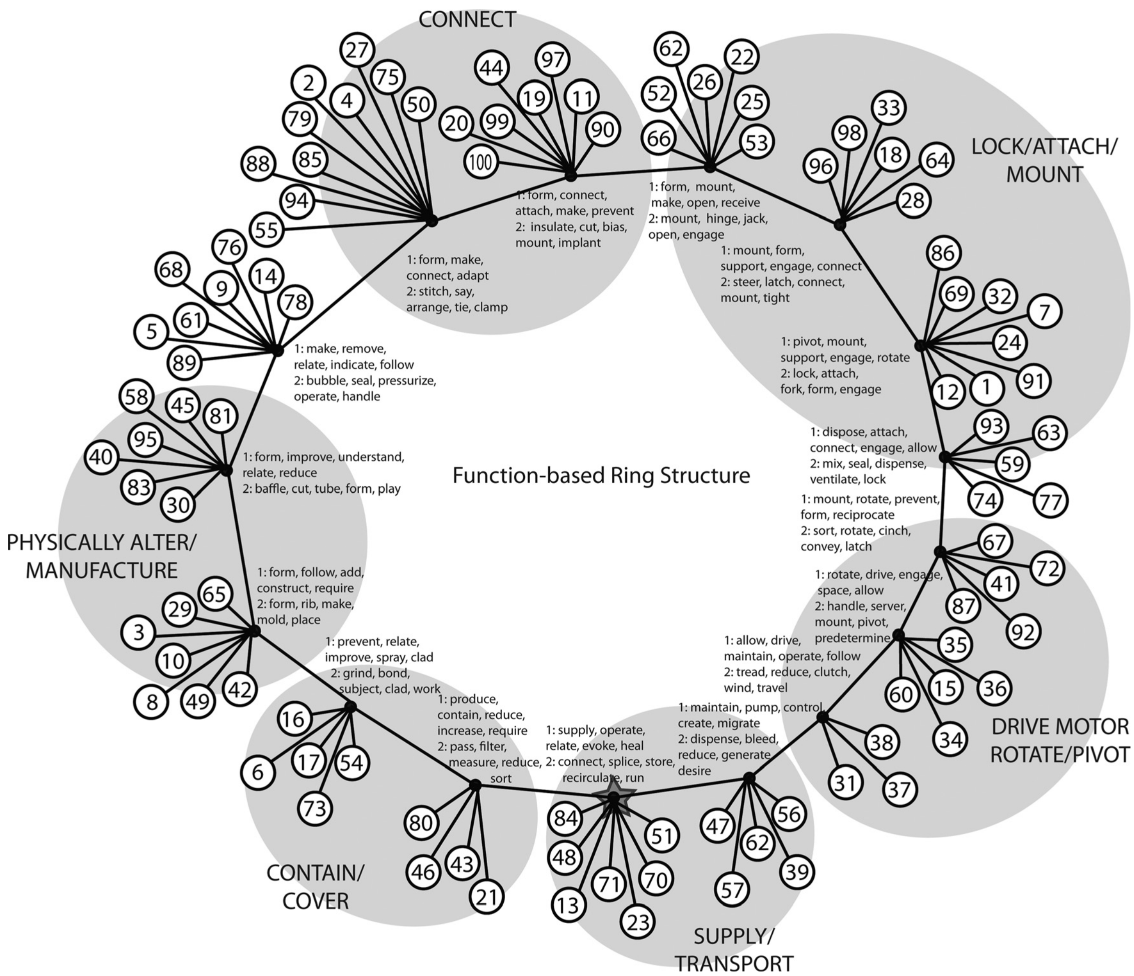

Fig. 4 Third best function based structure, ring showing regions of functionality

catheter, in the "medical" region, as they are both technologies related to the field of medicine. This association of patents is one that would likely be made by a human, grouping based on field of application. Another example is the grouping of patent 60, interactive toy (a Furby), and patent 69, bale handling apparatus, which both are mechanical devices, powered by motors. In this example, what is interesting is that the field of application is different, but the components or features of the designs and technologies are overlapping. This is evidence that cross-industry associations can happen in the surface based structures as well—indicating that both surface based and function based structures may be useful as the basis for computational design-by-analogy aids.

Figure 4 shows the regions extracted from the function based ring structure, some of which overlap with the hierarchy regions, and some of which do not. This is interesting, as it implies that different insights can be extracted from the different form types—making it valuable to not only examine the best structure, but the others as well. The regions in the function based ring structure are connect, lock/attach/mount, drive motor/rotate/pivot, supply/transport, contain/cover, and physically alter/manufacture. For example, while patents 25, a horseshoe, and patent 19, Attachments in model airplanes, were in the same region of "Secure/Attach/Mount" in the hierarchy structure, patent 19 is now in a region labeled "Connect." Patent 19, Attachments in model airplanes, is now in the ring structure associated with patents like patent 27, wound site management and wound closure device—where the analogous functionality is connecting parts of airplanes together, and connecting areas of human/animal tissue together. Another example is patent 84, smoke generating apparatus, and patent 46, method and apparatus for disposing of waste material—which are grouped in the "Connect" region together in the hierarchy structure, likely due to the connections required for material (smoke or waste) flow, but which are placed in separate regions in the ring structure of "Supply/transport" (patent 84) and "Contain/cover" (patent 46). The focus is shifted to the functionality of transporting the smoke or supplying the air for intake to the apparatus in patent 84, and the containment of the waste in patent 46. These insights demonstrate that different insights into functional relationships between patents can emerge by examining different forms of structures.

Finally, Fig. 5 displays the regions that emerge in the third best surface based structure, a ring. The regions for this structure are: motion/mechanical, fluid flow/control, medical, structure, and container. Many of these regions overlap with those found in the hierarchy structure in Fig. 3, though some do not. Patent 16, face mask, for example, is found in the "Medical" region of the ring structure, but found in the "Container" region of the hierarchy structure—both of which are sensible places to find this patent, but which contextualize it differently in terms of aspects of the patent to attend to. Patent 64, wheeled cart with removable skis, is found in the "Vehicle" region of the hierarchy structure, while found in the "Container" region of the ring structure—the first classifying it based on its transportation capabilities, while the second is classifying it based on its ability to hold or contain materials/goods. As seen in the discussion of Fig. 4, different insights can emerge as a result of examining the different form.

3.2 Type of Structure: Hierarchy versus Ring. As evidenced by the examples discussed in Sec. 3.1, there are different insights that emerge from representing the same patent space in

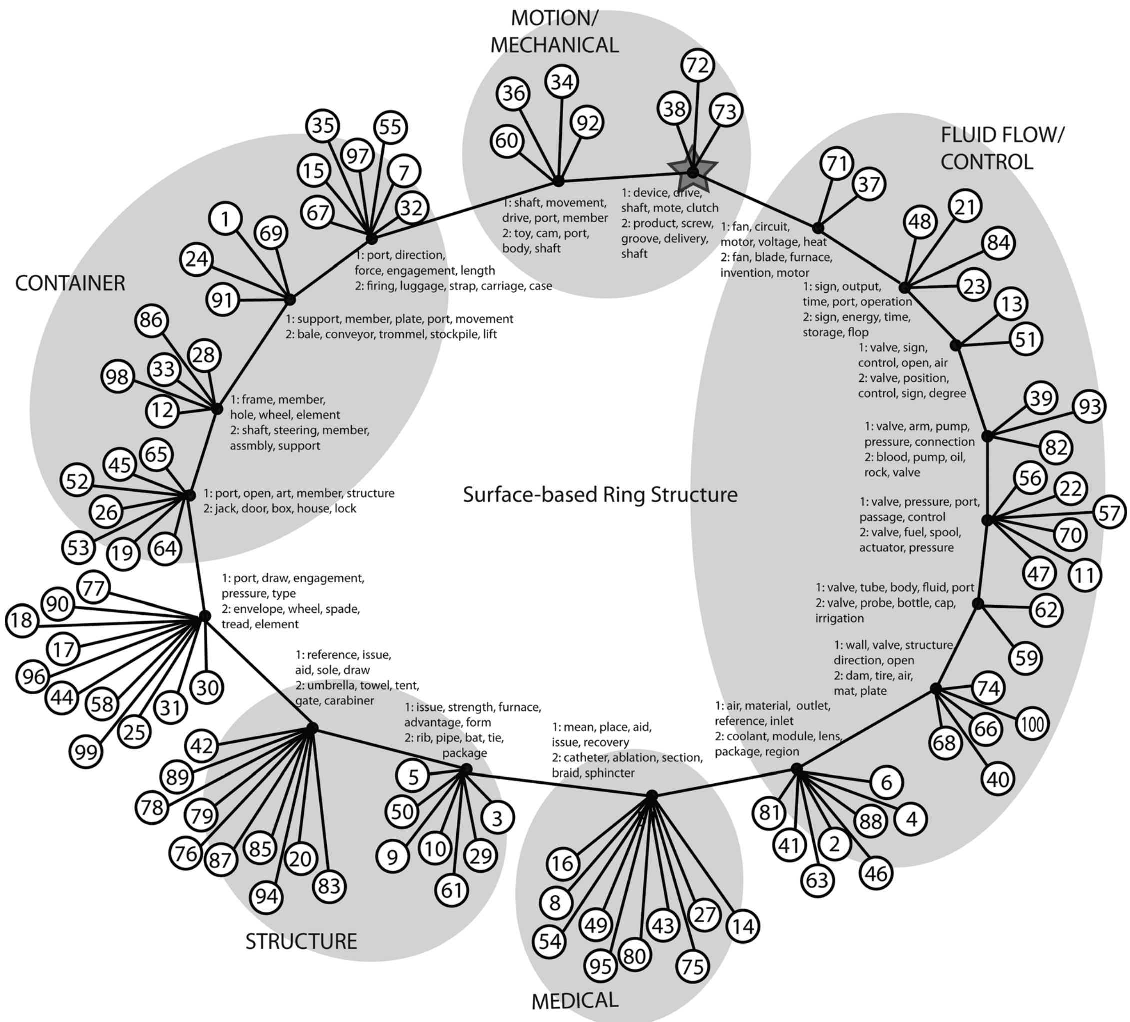

Fig. 5 Third best surface based structure, ring showing regions of surface content

different ways. This is one of the advantages of employing the Kemp and Tenenbaum algorithm, as it generates eight different representations of the space, along with scores corresponding to the order of quality of fit to the data. Rerepresenting a space with different regions or clustering can help a designer examine different functions or surface features of patents that can be attended to, which could lead to facilitation of cross-domain transfer of knowledge and access to far-field analogies through previously unconsidered connections between patents. For the purposes of stimulating design, multiple representations have been shown to be beneficial to the design process [10].

In addition, the forms explored in the algorithm and their generative graph grammars have different levels of restrictedness. The ring, for example, forces the space to be a connected continuum, which leads to a different arrangement of patents and clusters than might emerge in a hierarchical structure that allows for freer branching and sub-branching. This could actually be beneficial, as it is part of what may actually force the patents and clusters to be arranged in a new way, supplying a new representation to a designer that may be examining the structures.

3.3 Function Based Content versus Surface Based Content. The use of just function based content, the verbs, and just surface based content, the nouns, has proven to be an effective way of extracting different representations of the space that are based on different perspectives of the patents. It is clear that the function based structures, Figs. 2 and 4, and the examples called out in Sec. 3.1, that the patents are arranged and grouped based on related functionality. These relationships are easily understood in an efficient way by examining the cluster labels, as well as the region labels. The surface based structures, Figs. 3 and 5, and the examples called out in Sec. 3.1, show that surface based input data yield structures, regions and clusters of patents that communicate surface similarity, whether that is common field of application or common componentry among the clustered patents.

With respect to using these structures for design-by-analogy purposes, it would be interesting to perhaps combine the surface based structure with the function based structure in practice. If a cross-domain analogy was discovered through functional similarity by exploring the function structure, the area of application could be further explored by then looking to the surface based structure. Further analogical transfer could be achieved through the exploration of the surface features that are associated with the desired functionality—switching between the function based and surface based structure by discovering a patent or design of interest in one structure, and locating it in the other structure.

3.4 Design Problem in the Structure. In the content shown in Figs. 2–5, the function based and surface based hierarchy and ring structures show a continuity of functionality and surface content in the space, which is what enabled regions to be extracted. This continuity within the structure is exciting, as it implies that a designer seeking inspiration can traverse the structure in a meaningful way. As described in Sec. 2.4, the design problem was placed in the LSA space to find a "starting point," indicated by a red dot in all figures of structures in this paper. Generally speaking, the cluster selected as the starting point for all structures was one in which mechanical devices or motion was involved in the patents, and many of which were related to energy. Some examples are patent 82, torque limiting control, patent 72, single drive, multiscrew sorter with pusher means, patent 70, fuel injection system for linear engines, patent 38, synchronous drive pin clutch, patent 48, solar energy thermally powered electrical generating system, among others. This set of patents, while not directly solving the design problem at hand, does have potential for inspiring designs for capturing energy from human motion, whether through examining possible mechanisms to repurpose, energy sources to augment human energy with, or even possibly ways to store the energy once generated. From that starting point in each structure, a designer could branch out to explore other clusters nearby within the "mechanical" or "motion" region, or examine other desired subfunctionalities of the design problem or solution.

We anticipate that this type of guided exploration of the patent space will save the designer time, as compared to searching the USPTO website or other sources of analogical inspiration, and also contribute novel, inspirational, and relevant external stimuli. This methodology can, and in future work will, be applied to spaces including a much larger quantity of patents. As the patent space grows and the algorithm is run on much larger data sets, we expect the structures to become more complex and rich. More patents will allow for a smoother continuity of functionality over the structure, as there will be more fodder for transitions between nodes. We also expect that the clusters will become larger as more patents with more in common enter the space. The algorithm is currently run to favor simpler structures, or rather structures with fewer nodes; so as the patent space grows, this leads to larger clusters as opposed to smaller clusters in larger quantities.

3.5 Opportunities and Limitations. While the structures of design data in this work show great promise for being the basis of an automated design-by-analogy facilitation tool, there is still much work to be done to turn it into a usable, marketable computational design tool. First, the method must be scaled up significantly to include a much larger set of patents in order to achieve a reasonable probability of encountering very high quality analogical stimuli with respect to the design problem. This requires streamlining of the code, and a way of automatically prefiltering the patents if necessary to include only those of interest. Second, a way of quantitatively evaluating the quality of the cluster labels is needed. Currently, the labels are validated by expert opinion or manual cross-checking with patent content, an inefficient and subjective process. The scoring of the quality of the labels would enable a more streamlined process for improving them, in that they could be quickly and objectively evaluated, rerun, and reevaluated in an iterative way that is not currently easy to accomplish. In addition, visualization is a crucial element to making this methodology usable. An engineering designer needs a way to navigate the space with a smooth and intuitive interface, allowing them to zoom in and out, expand sections of interest in the structures, digest and analyze the actual content of a patent while navigating the structure, and visualizing thousands of patents or just a few patents at once, both in a meaningful, informative, and ideally inspiring way. This element is mostly a matter of developing a user interface, though testing it to maximize usability with user feedback would be necessary. Ultimately, the key beneficiaries of this work are engineering designers undertaking new and innovative product design; when compared to what most average designers use to research patents—internet searches, this tool has the potential to save them time and give them more inspiring and fruitful results.

In terms of the usefulness of such a tool, and the associated, underlying identification of structural form, recent research considers cognitive science studies of the proposed approach [80–82]. Exemplar results from these studies indicate that if the goal of conceptual ideation is to ultimately generate and develop a concept that is high quality and novel, then analogizing over relatively far-field, less-common examples is an effective way to do this. There are also implications for the design of tools and methods to support design-by-analogy. Computational tools that are able to define and compute functional and surface similarity between items in a design space in a principled manner relative to a current design problem would hold excellent potential as aids for inspiration. These tools might be able to maximize the potential benefits of analogies by retrieving and delivering to the designer in a timely manner surface dissimilar analogies and potentially even analogies that do not necessarily provide direct solutions to the target problem. If these systems are able to give priority to analogies that are relatively unusual or infrequently encountered, the potential for inspiration might be even higher

4 Conclusions

The goal of this work is to create a foundation for a computational design tool that allows designers to have automatic access to analogical stimuli from a design repository. We have presented results of a methodology that combines latent semantic analysis preprocessing and postprocessing of the data with Kemp and Tenenbaum's Bayesian model for discovering structural form [36]. We have shown that this methodology, as applied to a repository of random designs from the U.S. patent database, preprocessed to contain only function based content and only surface based content, has promising implications for the development of a computational design tool. The method has proven to produce diverse structures of patent data that lead to insights regarding the functional or surface relatedness of individual or groups of patents. The algorithm serves to uncover structure that could produce innovative thought and previously unconsidered relationships, ultimately leading to the potential identification of useful analogical stimuli in design practice.

Acknowledgment

The authors would like to thank Dr. Christian Schunn for his comments and insights on this work, and Dr. Jeremy Murphy for his contribution and insights regarding patent text processing. This work is supported by the National Science Foundation, under grant CMMI0855326.

Appendix A: 100 Randomly Selected Patents Used

Patent index number	U.S. patent number	Patent title
1	5,819,950	Portable trommel
2	4,506,651	Wood stove
3	5,239,707	Method of manufacturing apparatus for restraining a necktie
4	4,535,756	Solar collectors
5	4,304,332	Package
6	6,612,806	Turbo-engine with an array of wall elements that can be cooled and method for cooling an array of wall elements
7	5,762,169	Retractable auxiliary luggage case attachment and security tether mechanism and method
8	6,716,115	Thread wound golf ball
9	4,678,083	Intrusion indicating shield for consumer products
10	4,813,672	Batters' box
11	4,649,970	Magnetically actuated vapor valve
12	6,481,735	Apparatus for carrying a load behind a bicycle
13	4,380,233	Control device for an artificial respirator
14	6,991,656	Method and apparatus for performing a minimally invasive total hip arthroplasty
15	7,059,508	Surgical stapling instrument incorporating an uneven multistroke firing mechanism having a rotary transmission
16	4,488,547	Face mask

Appendix. ***Continued***

Patent index number	U.S. patent number	Patent title
17	5,228,241	Method and machine for grinding
18	4,379,706	Slidable-type constant velocity universal joint
19	4,233,773	Attachments in model airplanes
20	5,416,955	Trigger-closing carabiner
21	7,215,986	Signal processing apparatus
22	4,705,065	Safety relief system for control or vent valves
23	4,432,481	Splice-in-register control
24	4,203,505	Car hoist
25	3,970,149	Horseshoe
26	5,590,608	Lockable lock box mounting assembly and method
27	6,348,064	Wound site management and wound closure device
28	5,062,652	Sulky
29	5,768,928	Method of making an hydraulically efficient ribbed pipe
30	6,044,919	Rotary spade drill arrangement
31	5,921,843	Remote controlled toy vehicle
32	4,483,066	Apparatus for locking fasteners
33	6,776,447	Vehicle roof with a top which is movable between closed and open positions
34	4,913,681	Shock absorbing rotary gear coupling
35	5,964,159	Inclined or vertical lift
36	6,505,991	Self-centering shaft adapter
37	6,616,409	Method of designing an Impeller blade
38	6,769,593	Synchronous drive pin clutch
39	6,782,855	Valve train and method for reducing oil flow to deactivated engine valves
40	4,251,075	Maze game apparatus
41	5,528,878	Automated apparatus and method for consolidating products for packaging
42	5,819,391	Surface fastener and method of manufacturing the same
43	5,842,652	Waste recyclable processing mechanism
44	4,407,173	Device for removing insulation from an insulated conductor
45	4,230,228	Pin type solid butt rotary coupler
46	5,305,697	Method and apparatus for disposing of waste material
47	4,241,749	Pressure compensating valve
48	4,876,854	Solar energy thermally powered electrical generating system
49	6,143,013	High performance braided catheter
50	6,186,701	Elongate flexible container
51	5,931,180	Electropneumatic positioner
52	7,083,469	Modular mounting sleeve for jack
53	5,437,133	Grille fastener assembly
54	6,439,451	Method of making aluminum alloy plate for bearing
55	7,225,722	Linear drive
56	5,265,643	Constant flow rate control valve with low pressure drop start
57	5,984,148	Self-cleaning pressure relief and bypass valve, dispensing apparatus and method
58	5,375,948	Cutting insert for cutting and grooving tools
59	6,367,521	Gravity feed fluid dispensing valve
60	6,497,607	Interactive toy
61	4,853,977	Patient garment
62	5,993,410	Adjustable probe
63	4,223,996	Apparatus for mixing solid and liquid constituents of mortar or the like
64	4,589,668	Wheeled cart with removable skis
65	3,962,735	Movable bulkhead with guiding and overcanting prevention means
66	4,124,051	Shock absorbing wheel hub
67	7,175,212	Latch having releasable cinching mechanism
68	4,984,583	Air bubbling mats for therapeutically agitating bath water

Appendix. ***Continued***

Patent index number	U.S. patent number	Patent title
69	4,259,034	Bale handling apparatus
70	6,634,325	Fuel injection system for linear engines
71	4,123,000	Method of starting a hot air furnace
72	5,909,815	Single drive, multiscrew sorter with pusher means
73	3,975,130	Installation for producing glassceramic tiles
74	4,103,708	Ventilated poppet damper
75	3,964,473	Bone prosthesis
76	4,705,064	Safety seal for an operating lever
77	6,142,689	Envelope leveler for printer feeder
78	5,273,173	Screw top
79	5,438,724	Method for using plastic fasteners for shoe-lasting applications
80	6,974,456	Method to treat gastric reflux via the detection and ablation of gastro-esophageal nerves and receptors
81	4,867,134	Fluid-heating solar collector
82	3,941,514	Torque limiting control
83	6,109,282	Self-erecting loop structure
84	4,303,397	Smoke generating apparatus
85	5,899,571	Beach towel, tote bag and beach umbrella system
86	6,234,452	Hand operable motorcycle stand
87	4,841,621	Shaft adjuster
88	4,142,679	Building heating system
89	6,634,044	Compact stretcher
90	4,270,310	Support device for an upstanding plant support rod in a plant pot
91	5,423,097	Emergency drop fowler and gatch
92	5,572,898	Modular die transfer system
93	3,938,909	Single needle alternating flow blood pump system
94	5,647,066	Safety helmet visor
95	6,119,041	Apparatus and method for linear lesion ablation
96	4,484,762	Ski binding and boot
97	4,762,262	Side-fed stapler
98	6,164,698	Steering device for automobiles
99	6,062,856	Dental implant hole guide extension
100	4,739,727	Animal waterer

References

[1] Ullman, D. G., 2003, *The Mechanical Design Process*, 3rd ed., McGraw-Hill Companies, New York, NY.

[2] "Cross-Industry Innovation," Retrieved From Business Process Optimization, European Ceo: http://Www.Zuehlke.com/Fileadmin/Pdf/Fachartikel/124_Cd_European_Ceo_Cross_Innovation.Pdf, April 9, 2009.

[3] Basantani, M., 2009, "The Maglev: The Super-Powered Magnetic Wind Turbine," Retrieved From Inhabitat Energy: http://Www.Inhabitat.com/2007/11/26/Super-Powered-Magnetic-Wind-Turbine-Maglev/, April 9, 2009.

[4] Linder, P., "Leveraging Novel Mems Technologies For Next Generation Photovoltaic Applications," http://www.Semiconductor.Net/Article/Ca6576216.Html, April 11, 2009.

[5] Herstatt, C., and Kalogerakis, K., 2004, "How to Use Analogies for Breakthrough Innovations," Working Papers/Technologie-Und Innovationsmanagement 24, http://Biblioteca.Universia.Net/Html_Bura/Ficha/Params/Id/39800023.Html, April 11, 2009.

[6] Bloom, P., 1996, "Intention, History, and Artifact Concepts," Cognition, **60**, pp. 1–29.

[7] Defeyter, M. A., and German, T. P., 2003, "Acquiring an Understanding of Design: Evidence From Children's Insight Problem Solving," Cognition, **89**, pp. 133–155.

[8] Dahl, D. W., and Moreau, P., 2002, "The Influence and Value of Analogical Thinking During New Product Ideation," J. Mark. Res., **39**, pp. 47–60.

[9] Goldschmidt, G., and Smolkov, M., 2006, "Variances in the Impact of Visual Stimuli on Design Problem Solving Performance," Des. Stud., **27**, pp. 549–569.

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

[11] Christensen, B. T., and Schunn, C. D., 2005, "Spontaneous Access and Analogical Incubation Effects," Creat. Res. J., **17**, pp. 207–220.

[12] Christensen, B. T., and Schunn, C. D., 2007, "The Relationship of Analogical Distance to Analogical Function and Preinventive Structure: The Case of Engineering Design," Mem. Cognit., **35**, pp. 29–38.
[13] Tseng, I., Moss, J., Cagan, J., and Kotovsky, K., 2008, "The Role of Timing and Analogical Similarity in the Stimulation of Idea Generation in Design," Des. Stud., **29**, pp. 203–221.
[14] Moss, J., Kotovksy, K., and Cagan, J., 2007, "The Influence of Open Goals in the Acquisition of Problem Relevant Information," J. Exp. Psychol. Learn. Mem. Cogn., **33**, pp. 876–891.
[15] Moss, J., Cagan, J., and Kotovsky, K., 2007, "Design Ideas and Impasses: The Role of Open Goals.," Presented at the Proceedings of the 16th International Conference on Engineering Design.
[16] Jansson, D. G., and Smith, S. M., 1991, "Design Fixation," Des. Stud., **12**, pp. 3–11.
[17] Smith, S. M., and Blankenship, S. E., 1991, "Incubation and the Persistence of Fixation in Problem Solving," Am. J. Psychol., **104**, pp. 61–87.
[18] Purcell, A. T., and Gero, J. S., 1996, "Design and Other Types of Fixation," Des. Stud., **17**, pp. 363–383.
[19] Chrysikou, E. G., and Weisberg, R. W., 2005, "Following the Wrong Footsteps: Fixation Effects of Pictorial Examples in a Design Problem Solving Task," J. Exp. Psychol. Learn. Mem. Cogn., **31**, pp. 1134–1148.
[20] Marsh, R. L., Ward, T. B., and Landau, J. D., 1999, "The Inadvertent Use of Prior Knowledge in a Generative Cognitive Task," Mem. Cogn., **27**, pp. 94–105.
[21] Smith, S. M., Ward, T. B., and Schumacher, J. S., 1993, "Constraining Effects of Examples in a Creative Generation Task," Mem. Cogn., **21**, pp. 837–845.
[22] Stone, R., and Wood, K. L., 2000, "Development of a Functional Basis for Design," J. Mech. Des., **122**, p. 359.
[23] Cheong, H., Shu, L. H., Stone, R., and Wood, K. L., 2008, "Translating Terms of the Functional Basis into Biologically Meaningful Keywords," Presented at the Proceedings of the ASME International Design Engineering Technical Conference, New York, NY.
[24] Altshuller, G. S., and Shapiro, R. B., 1956, "О Психологии Изобретателского творчества (On the Psychology of Inventive Creation) (In Russian)," *Вопросы Психологии* (The Psychological Issues), **6**, pp. 37–39.
[25] Zhang, R., Cha, J., and Lu, Y., 2007, "A Conceptual Design Model Using Axiomatic Design, Functional Basis and Triz," Presented at the Proceedings of the 2007 IEEE.
[26] Souili, A., Cavallucci, D., Rousselot, F., and Zanni, C., 2011, "Starting From Patents to Find Inputs to the Problem Graph Model of Idm-Triz," Presented at the Triz Future 2011, Dublin, Ireland.
[27] Souili, A., and Cavallucci, D., 2012, "Toward an Automatic Extraction of IDM Concepts From Patents," Presented at the Cirp Design.
[28] Cascini, G., and Russo, D., 2007, "Computer-Aided Analysis of Patents and Search for Triz Contradictions," Int. J. Prod. Dev., **4**, pp. 52–67.
[29] Chakrabarti, A. K., Dror, I., and Nopphdol, E., 1993, "Interorganizational Transfer of Knowledge: An Analysis of Patent Citations of a Defense Firm," IEEE Trans. Eng. Manage., **40**, pp. 91–94.
[30] Indukuri, K. V., Ambekar, A. A., and Sureka, A., 2007, "Similarity Analysis of Patent Claims Using Natural Language Processing Techniques," Presented at the International Conference on Computational Intelligence and Multimedia Applications.
[31] Kasravi, C., and Risov, M., 2007, "Patent Mining—Discovery of Business Value From Patent Repositories," Presented at the Proceedings of the 40th Hawaii International Conference on System Sciences.
[32] Moehrle, M., and Geritz, A., 2004, "Developing Acquisition Strategies Based on Patent Maps," Presented at the 13th Iamot, Washington, DC.
[33] Bohm, M. R., Vucovich, J. P., and Stone, R. B., 2005, "Capturing Creativity: Using a Design Repository to Drive Concept Innovation," Presented at the Proceedings of DETC2005, Long Beach, California, Paper No. DETC05/Cie-85105.
[34] Szykman, S., Sriram, R. D., Bochenek, C., Racz, J. W., and Senfaute, J., 2000, "Design Repositories: Engineering Design's New Knowledge Base," IEEE Intellegent Systems, pp. 48–55.
[35] Koch, S., Bosch, H., Giereth, M., and Ertl, T., 2009, "Iterative Integration of Visual Insights During Patent Search and Analysis," Presented at the IEEE Symposium on Visual Analytics Science and Technology, Atlantic City, NJ.
[36] Mukherjea, S., Bhuvan, B., and Kankar, P., 2005, "Information Retrieval and Knowledge Discovery Utilizing a Biomedical Patent Semantic Web," IEEE Trans. Knowl. Data Eng., **17**, pp. 1099–1110.
[37] Chakrabarti, S., Dom, B., Agrawal, R., and Raghavan, P., 1998, "Scalable Feature Selection, Classification and Signature Generation for Organizing Large Text Databases Into Hierarchical Topic Taxonomies," VLDB J., **7**, pp. 163–178.
[38] Duran-Novoa, R., Leon-Rovira, N., Aguayo-Tellez, H., and Said, D., 2011, "Inventive Problem Solving Based on Dialectical Negation, Using Evolutionary Algorithms and Triz Heuristics," Comput. Ind., **62**, pp. 437–445.
[39] Hernandez, N. V., Schmidt, L. C., and Okudan, G. E., 2012, "Systematic Ideation Effectiveness Study of Triz," Presented at the Asme IDETC/CIE, Chicago, IL.
[40] Hernandez, N. V., Schmidt, L. C., and Okudan, G. E., 2012, "Experimental Assessment of Triz Effectiveness in Idea Generation," Presented at the ASEE Annual Conference, San Antonio, TX.
[41] Krasnoslobodtsev, V., and Langevin, R., 2005, "Triz Application in Development of Climbing Robots," Presented at the First Triz Symposium, Japan.
[42] Liang, Y., Tan, R., and Ma, J., 2008, "Patent Analysis With Text Mining for Triz," Presented at the IEEE ICMIT, Bangkok, Thailand.
[43] Nakagawa, T., 2012, "Creative Problem-Solving Methodologies Triz/Usit: Overview of My 14 Years in Research, Education, and Promotion," Bulletin of the Cultural and Natural Sciences in Osaka Gakuin University, Vol. 64.
[44] Nix, A. A., Sherret, B., and Stone, R. B., 2011, "A Function Based Approach to Triz," Presented at the ASME IDETC/CIE, Washington, DC.
[45] Houssin, R., and Coulibaly, A., 2011, "An Approach to Solve Contradiction Problems for Safety Integration in Innovative Design Process," Comput. Ind., **62**, pp. 398–406.
[46] "Invention Machine Goldfire: Unleashing the Power of Research," available at http://Inventionmachine.com/Products-And-Services/Innovation-Software/Goldfire-Research/, February19, 2012.
[47] "Creax: Creativity for Innovation," available at http://www.creax.com, September 7, 2012.
[48] Verhaegen, P., D'hondt, J., Vandevenne, D., Dewulf, S., and Duflou, J. R., 2011, "Identifying Candidates for Design-by-Analogy," Comput. Ind., **62**, pp. 446–459.
[49] Goel, A., Bhatta, S., and Stroulia, E., 1997, "Kritik: An Early Case-Based Design System," *Issues and Applications of Case-Based Reasoning in Design*, M. Maher and P. Pu, eds., Erlbaum, Ed Mahwah, NJ, pp. 87–132.
[50] Bhatta, S., and Goel, A., 1996, "From Design Experiences to Generic Mechanisms: Model-Based Learning in Analogical Design," Aiedam, **10**, pp. 131–136.
[51] Chiu, I., and Shu, L. H., 2005, "Bridging Cross-Domain Terminology for Biomimetic Design," ASME IDETC/CIE, Long Beach, CA.
[52] Vincent, J. F. V., Bogatyreva, O. A., Bogatyreva, N. R., Bowyer, A., and Pahl, A. K., 2006, "Biomimetics: Its Practice and Theory," J. R. Soc. Interface, **3**, pp. 471–482.
[53] Gick, M. L., and Holyoak, K. J., 1980, "Analogical Problem Solving," Cognit. Psychol., **12**, pp. 306–355.
[54] Forbus, K. D., Gentner, D., and Law, K., 1994, "Mac/Fac: A Model of Similarity-Based Retrieval," Cognit. Sci., **19**, pp. 141–205.
[55] Griffiths, T. L., Kemp, C., and Tenenbaum, J. B., 2008, "Bayesian Models of Cognition," *Cambridge Handbook of Computational Psychology*, R. Sun, ed., Cambridge University Press, New York, NY, pp. 59–100, DOI: 10.1016/0364-0213(95)90016-0.
[56] Jaynes, E. T., 2003, "Plausible Reasoning," *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge, Chap. 1.
[57] Kemp, C., and Tenenbaum, J. B., 2008, "The Discovery of Structural Form," Proc. Natl. Acad. Sci. U.S.A., **105**, pp. 10687–10692.
[58] Nosofsky, R. M., 1986, "Attention, Similarity, and the Identification-Categorization Relationship," J. Exp. Psychol. Gen., **115**, pp. 39–61.
[59] Ashby, F. G., and Maddox, W. T., 1992, "Complex Decision Rules in Categorization: Contrasting Novice and Experienced Performance," J. Exp. Psychol.: Hum. Percept. Perform., **18**, pp. 50–71.
[60] Anderson, J. R., 1991, "The Adaptive Nature of Human Categorization," Psychol. Rev., **98**, pp. 409–429.
[61] Kemp, C., Tenenbaum, J. B., Griffiths, T. L., Yamada, T., and Ueda, N., 2006, "Learning Systems of Concepts With an Infinite Relational Model," Presented at the Proceedings of the 21st National Conference on Artificial Intelligence.
[62] Kemp, C., and Tenenbaum, J. B., 2009, "Structured Statistical Models of Inductive Reasoning," Psychol. Rev., **116**, pp. 20–58.
[63] Griffiths, T. L., Steyvers, M., and Tenenbaum, J. B., 2007, "Topics in Semantic Representation," Psychol. Rev., **114**, pp. 211–244.
[64] Landauer, T. K., Foltz, P. W., and Laham, D., 1998, "An Introduction to Latent Semantic Analysis," Discourse Process., **25**, pp. 259–284.
[65] Foltz, P. W., Kintsch, W., and Landauer, T. K., 1998, "The Measurement of Textual Coherence With Latent Semantic Analysis," Discourse Process, **25**, pp. 285–307.
[66] Deerwester, S., Dumais, S. T., Furnas, G. W., and Landauer, T. K., 1990, "Indexing by Latent Semantic Analysis," J. Am. Soc. Inf. Sci., **41**, pp. 391–407.
[67] Shepard, R. N., 1980, "Multidimensional Scaling, Tree-Fitting, and Clustering," Science, **210**, pp. 390–398.
[68] Inhelder, B., and Piaget, J., 1969, *The Early Growth of Logic in the Child*, W W Norton & Company New York, NY.
[69] Huelsenbeck, J. P., and Ronquist, F., 2001, "Mrbayes: Bayesian Inference of Phylogenetic Trees," Bioinformatics, **17**, pp. 754–755.
[70] Fiske, A. P., 1992, "The Four Elementary Forms of Sociality: Framework For a Unified Theory of Social Relations," Psychol. Rev., **99**, pp. 689–723.
[71] Guttman, L., 1944, "A Basis for Scaling Qualitative Data," Am. Sociol. Rev., **9**, pp. 139–150.
[72] Bradley, R. A., and Terry, M. E., 1952, "Rank Aanalysis of Incomplete Block Designs. 1. The Method of Paired Comparisons," Biometrika, **39**, pp. 324–345.
[73] Guttman, L., 1954, "A New Approach to Factor Analysis: The Radex," *Mathematical Thinking in the Social Sciences*, P. F. Lazarsfeld, ed., Free Press, New York, NY, pp. 258–348.
[74] Wiggins, J. S., 1996, "An Informal History of the Interpersonal Circumplex Tradition," J. Per. Assess., **66**, pp. 217–233.
[75] Sneath, P. H., and Sokal, R. R., 1973, *Numerical Taxonomy: The Principles and Practice of Numerical Classification*, Freeman, San Francisco, CA.
[76] Collins, A. M., and Quillian, M. R., 1969, "Retrieval Time From Semantic Memory," J. Verbal Learn. Verbal Behav., **8**, pp. 240–247.
[77] Carroll, J. D., 1976, "Spatial, Nonspatial and Hybrid Models for Scaling," Psychometrika, **41**, pp. 439–463.
[78] Kohonen, T., 1997, *Self-Organizing Maps*, Springer, New York, NY.
[79] Kemp, C., and Tenenbaum, J., 2008, "The Discovery of Structural Form," PNAS, Supporting Information Appendix 10.1073/Pnas.080263110.

[80] Fu, K., 2012, "Discovering and Exploring Structure in Design Databases and Its Role in Stimulating Design by Analogy," Ph.D. dissertation, Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, PA.
[81] Fu, K., Chan, J., Cagan, J., Kotovsky, K., Schunn, C., and Wood, K., 2012, "The Meaning of "Near" and "Far": The Impact of Structuring Design Databases and the Effect of Distance of Analogy on Design Output," ASME J. Mech. Des. (in press).
[82] Chan, J., Fu, K., Schunn, C., Cagan, J., Wood, K., and Kotovsky, K., 2011, "On the Benefits and Pitfalls of Analogies for Innovative Design: Ideation Performance Based on Analogical Distance, Commonness, and Modality of Examples," J. Mech. Des., **133**, p. 081004.