A METHODOLOGY FOR DISCOVERING STRUCTURE IN DESIGN DATABASES

~~Author(s) – LEAVE BLANK IN FIRST SUBMISSION~~

ABSTRACT

Design by analogy, in which designers draw inspiration from cross-domain design solutions, is a promising methodology for product development. This work attempts to leverage the existing design solutions within a repository, combined with an exploration of inherent structural forms that can be discovered based on the content and similarity of that data, in order to gain useful insights into the nature of the design space. In this work, the approach will be applied to uncover structure in the U.S. patent database. Methodology for generating and analyzing these structures is presented here, in addition to insights from some initial results. These insights could generate fodder for stimulating design inspiration and innovation for engineering designers.

Keywords: design by analogy, latent semantic analysis, patent, computational design tool

1 INTRODUCTION

Design has become a ubiquitous force in many areas of society, from policy making to education, from services and experiences to tangible products. We are becoming acutely aware of the impact that good design can have on the success of anything that is being created. Good design can, among many other things, make us more efficient, happier, smarter, more responsible, and more comfortable, and can enhance our quality of life. In order to achieve these remarkable effects of good design, we need good designers. Good designers come about through inherent talent, the training they receive in design processes, and the external environment and information to which they are exposed. Here, we look to the aspects that we can control, i.e., the process that any designer can use in order to achieve good design and the external stimuli to which they are exposed. Among a multitude of methodologies and philosophies behind effective design processes, one proven approach to achieving innovative solutions is 'design-by-analogy' [1]. Design-by-analogy is a process in which designers are exposed to and make use of design solutions from other domains in order to gain inspiration or insight for the design problem at hand. Here, we focus on engineering design and the product development process, though design-by-analogy can be applied in any domain. Design-by-analogy is becoming more popular with designers in industry, whether it is by applying MEMS 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, 3, 4, 5]. Although this methodology has the potential to be incredibly fruitful in the engineering product design process, there lacks a practical, efficient, procedural way to find these meaningful analogies.

In looking for cross-domain design solutions, it is particularly convenient and advantageous that a large number of design solutions are documented in the United States patent database. Using this repository as a source of analogical inspiration is a logical choice. However, due to the size and complexity of the US patent database, it's extraordinarily 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, including TRIZ (theory of inventive problem solving) [6]. TRIZ uses heuristic rules (such as use of opposites) to help engineers overcome impasses in functional reasoning by searching through patents [6]. Tools like TRIZ 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. With a way to extract the interrelatedness and interconnectedness of patents in the space, 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.

Creating a more meaningful and insightful organization of patents is a large task simply due to the amount of information and data that is involved. In addition, it is desirable to organize that information in a way that is in sync with how humans think, categorize and organize information cognitively. Thus, we look to the field of computer science as applied to modeling human cognition for methodologies to achieve this goal. Bayesian models have been used for over 200 years to describe human cognition, most recently to model human inductive learning, semantic memory, causal learning and inference, and categorization, among many others [7].

Kemp and Tenenbaum used Bayesian inference as the foundation for an algorithm that discovers structural form in data [8]. By discovering the structure that is inherent to the data, they were able to extract meaningful information that was not readily accessible by simply looking at the raw data. For example, given a set of sixteen colors, the structural form discovered by the algorithm that best described the data was a ring – implying the color wheel that we are all familiar with from our elementary or primary school days. Given data regarding votes of Supreme Court justices, the best structural form discovered was a chain, organizing the justices from liberal to conservative. Given data regarding features that a set of animals have, a tree structure was discovered to be the best, not dissimilar to the biological classification scheme recognized by Linnaeus [8]. Imagine looking at a large matrix of any of these data sets and trying to extract meaningful insights about the relationships between the entities (i.e., colors, animals, justices, etc.) described. As the data set becomes larger, the ability to synthesize that data with the human mind alone becomes an intractable problem. Now consider the immense amount of data that is the U.S. patent database, and attempting as an engineering designer to synthesize that information into inspiring cross-domain analogies. We hypothesize that by applying Kemp and Tenenbaum's algorithm to patents, structural form of the patent space can be discovered, which could uncover insights about how these patents are meaningfully interrelated. Additionally, the structure finding algorithm is combined with a post-processing of the structures using latent semantic analysis (LSA), which allows for the structures to be labeled with informative text based on the content of the patents and their connections. We hypothesize that the combination of the algorithm with LSA will allow for the structures to be visually interpretable and informative without additional information about the patents, which is a promising step toward a structure-based design repository analogical stimulation tool.

2 BACKGROUND

2.1 Cognitive Mechanisms in Design

The use of analogy in design has been studied a great deal in the fields of cognitive science and engineering design. Work has been done to understand how the introduction of analogies affects the ideation process and outcomes [9, 10, 11, 12], with some studies specifically examining how the introduction of analogies with different levels of applicability to the design problem affects individual designers [13, 14]. It has been shown that if designers 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 [15, 16]; this open goal effect is achieved by giving designers supplemental valuable information, or hints consisting of distant or unobvious information, only after solving has already begun.

Tseng et al. found that giving individuals information that was analogous but distantly related to the design problem caused them to produce more solutions with a wider range of solution types and 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 [14]. A potential negative effect of introducing analogical information or examples that has been extensively explored is design fixation [17, 18, 19, 20], or the "blind adherence to a set of ideas or concepts limiting the output of conceptual design" [17]. Jansson & Smith showed that introducing examples can cause designers to generate solutions that mimic the examples, to the point of violating the design problem objectives [17]. 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 [21, 22]. With the work proposed here, the expectation is to uncover inspirational analogical information that could be useful to the designer if introduced at the appropriate point in ideation, in the most helpful format and under the best conditions.

2.2 Design Tools

Creating computational tools to aid designers during the design process is a popular end goal of research in the area of engineering design theory and methodology. 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 [23], which has been extended and adapted a great deal, one example of which is a biological functional basis [24]. This functional basis and language of design work will be an important aspect of the work proposed here, as it will allow for the exploration of functional interrelatedness of patents as compared to surface interrelatedness.

The use of patents as design aids is not a new area of research either. TRIZ, a heuristic based tool that helps designers overcome impasses during ideation, hinges on the idea that a solution to a design problem already exists in the patent literature, but perhaps in another field of application [6]. TRIZ and functional basis have been combined to create an axiomatic conceptual design model [25]. Methods have been developed to find the interrelatedness between technologies based on patent citation data, and the benefits of tapping into the technology knowledge base created by competitors within a particular design field [26]. Syntactic similarity between patent claims has been explored for the purpose of aiding in patent infringement research [27]. Patent repository tools and patent mining have been used to ascertain potential future market trends, recognize prolific inventors, etc., 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. [28]. 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 [29].

Mukherjea et al. created a BioMedical Patent Semantic Web, 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 includes the abstract of the patents being examined. In addition, the webs were not generated using a Bayesian inference approach [30].

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 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 [31]. The important difference between the work of Koch et al. and the work proposed is that the structures within the PatViz tool are either predefined or user-defined classification schemes, while the proposed 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.

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 [32]. 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.

2.3 Bayes and Discovering Structure

Finding the structure in data is not an easy or new problem, but has the potential to yield valuable insights if successful. Examples include Linneaus’ discovery that living organisms are best described by a tree structure, or Mendeleev’s finding of the periodic structure of the elements [8]. 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 [33, 34, 35, 36, 37]. Methods such as Latent Dirichlet Analysis and Latent Semantic Analysis have been used to categorize documents based on the text within them, extracting taxonomies and semantic similarity [38, 39, 40, 41].

Bayesian models have been used to describe human cognition for centuries [7]. 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 [42].

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. These structures originate from psychology literature [42] and appear in formal models in many different research efforts [43-55]. 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. [8]"

A structure is a particular instantiation of a form. To be clear, a graph of data 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 [8].

$$P(S, F|D) \propto P(D|S)\, P(S|F)\, P(F). \qquad (1)$$

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 which divides the right hand side of Equation 1, 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.

2.3 Latent Semantic Analysis

Latent semantic analysis (LSA) is a computational text analysis tool that extracts contextual similarity of documents and words [39, 40, 41]. LSA is comprised of four main steps.

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.
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 can be performed by altering the S matrix to only contain the top n values along the diagonal, which can lead to better results in analyses with large corpora. However, due to the small size of the corpora used here (10 patents are used to illustrated the method in this paper), 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 same, a value of 1 would be output for that cosine similarity [39, 40, 41].

LSA is used in this work to generate "feature" data for input into the structural form discovery algorithm. In addition, in order to be able to analyze the meaning of the connections between patents in the structural forms that are output, a characterization of each patent or connection between patents was devised. Latent Semantic Analysis was used to find the words in the LSA space that had the highest cosine similarity value to each patent. In addition, the set of words between any two set of patents in the space that had the highest summed cosine similarity to the two patents was found. These sets of words were used to label the graphs of the structural forms to allow them to be more easily interpretable. This methodology is described in detail in the following section.

3 METHODOLOGY

There are two parts to producing the structures presented in the results section. First, LSA was used to pre-process the patents, as well as to find the sets of words with the highest cosine similarity to each patent and connection between patents. Second, the algorithm devised by Kemp and Tenenbaum was used to discover structural forms in the patent data, using the output from LSA as input.

3.1 LSA Pre-processing

Given an initial set of patents, the abstract and description of the patents were first parsed from HTM text. A part-of-speech (POS) tagger was run on the abstract and description text file, and verbs, adverbs, adjectives and nouns were extracted separately from the POS tagged file. Repeated terms were included.

LSA was then run on the set of ten patents as described in section 2.3. A symmetric 10 x 10 cosine similarity matrix describing the similarity between the 10 patents resulted, and was used as input to the algorithm described in section 3.2. LSA was additionally used to find the set of words (here, the top 5) within the LSA space that had the highest cosine similarity values to each patent. These words were used to label the patent nodes on the structural forms in order to allow for a way to gain a better understanding of why the connections in the structures exist in each particular configuration with a text "snapshot" of the patents in structure. In addition, each pair of patents was analyzed to find the set of common words that had the highest summed cosine similarity values, which was used to characterize a connection between or adjacency of any two patents in a structure, lending further insight into the meaning of the connections.

3.2 Structural Form Discovery

The algorithm for discovering structural form as it was be applied to the LSA output patent "feature" data includes the following steps [8, 43]:

1. Pre-process the feature data *D* by shifting the mean of the matrix to zero. Calculate normalized covariance matrix for *D,* defined as $(1/m)DD^T$, where m is the number of features, or non-redundant non-trivial 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, were used as candidate forms.

4 RESULTS AND DISCUSSION

Kemp and Tenenbaum's algorithm attempts to fit a predefined group of structural forms to a given data set and returns the structural form with the best score, or the calculated log posterior probability that that particular data is described by that particular instantiation of that structural form. Ten patents were chosen randomly as the data for this work. The algorithm can accommodate a much larger set, but was kept small here to illustrate this new approach to uncover analogies, and for simplicity of analysis of the results to articulate the insights. The structural form that best fit the input set of ten patents was the ring structure. The patents used for this work are listed in Table 1.

Table 1. Patent Numbers and Corresponding Patent Titles for Ten Random Patents Used

Patent Number	Patent Title
3936685	Rotor stack for a squirrel-cage, sliding rotor type motor
4001224	4-Substituted-2, 3-dihydro-1-benzoxepin-3, 5-diones and tautomers
4746850	Start-up system for a synchronous motor drive
5578512	Power MESFET structure and fabrication process with high breakdown voltage and enhanced source to drain current
5655489	Valve cover
6111067	A-82846-type glycopeptide antibiotics
6230449	Support attachment for a post
6501034	integrated connector device including a cable reel and Combined switch systems
6791040	Locking assembly for an electrical switching apparatus
7158843	Modular software definable pre-amplifier

Figure 1 shows the ring structure of the ten patents, with labels below each patent node comprised of the five most similar words from the LSA space to each patent. As stated in the methodology section, these five most similar words were selected based on the cosine similarity calculation performed in LSA. Note that the structure in Figure 1 has multiple patents clustered together on a single node, and those nodes then lie on the ring form.

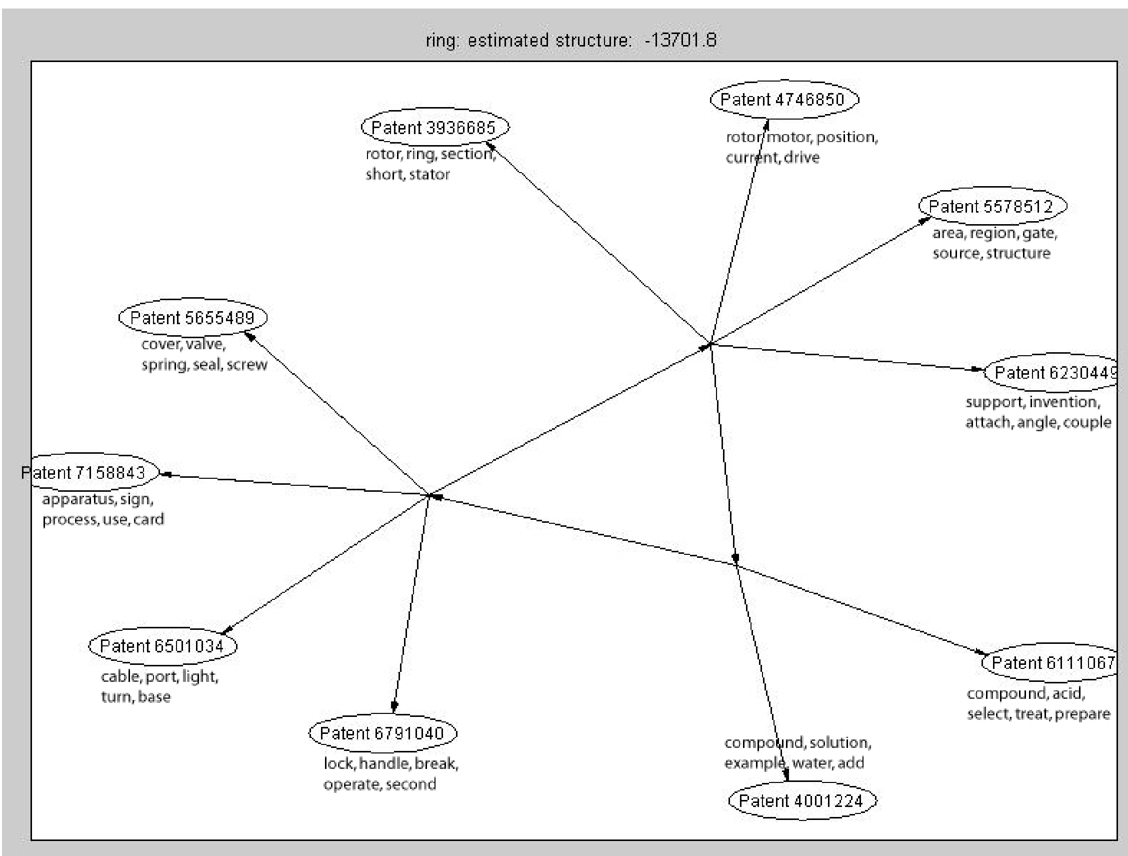

Figure 1. Ring Structure of Ten Random Patents,
Labeled with Five Most Similar Words to Each Patent as Derived from LSA

First, it is exciting to see that the top five words labels are extremely helpful in being able to immediately interpret the gist of each patent, as well as in being able to make some guesses about why certain patents are clustered together or connected. For example, the patents in the upper right corner of Figure 1 all have mechanical components, or descriptions of mechanical components involved. By looking at the corresponding titles in Table 1, we can verify that two out of four of the patents in that cluster have to do with motors, and the other two have to do with some sort of structure. The two patents clustered in the lower right of Figure 1 seem to both be related to chemistry based on the top five words for both of them. Table 1 allows for verification that indeed those two patents are for two particular chemicals. These meaningful clusterings not only verify that the Kemp and Tenenbaum's algorithm is yielding significant results, but also that the use of LSA to find text to display on the structures is beneficial, and can allow for a much more streamlined and efficient analysis of structural forms of patents.

Another scheme was devised for allowing the structures to be easily interpreted visually, which was finding the top five words that are jointly most common to any two patents. This labeling scheme led to quite different results than those presented in Figure 1. Figure 2 shows the same underlying ring structure that was used in Figure 1, with the top five most similar words to each connected patent pair outlined in red with connecting lines to the relevant pair. It's interesting, for example, that the pair in the lower right, which have been established as chemical patents, have words most in common that might seem to have less to do with chemistry. One hypothesis is that these words have to do more with necessary functionalities involved in the manufacture of new chemicals, as opposed to more surface similarity that seemed to emerge in Figure 1.

In addition, notice that the lower left cluster of patents in Figure 2 now includes words in common that were describing the application of patents in the upper right in Figure 1, with words like ring, rotor, motor, stator connecting many pairs. Another hypothetical explanation is that these connecting sets of words point to a common component or set of features in the cluster of patents that transcends area of application, perhaps leading to the first steps of cross industry application of technology and innovation.

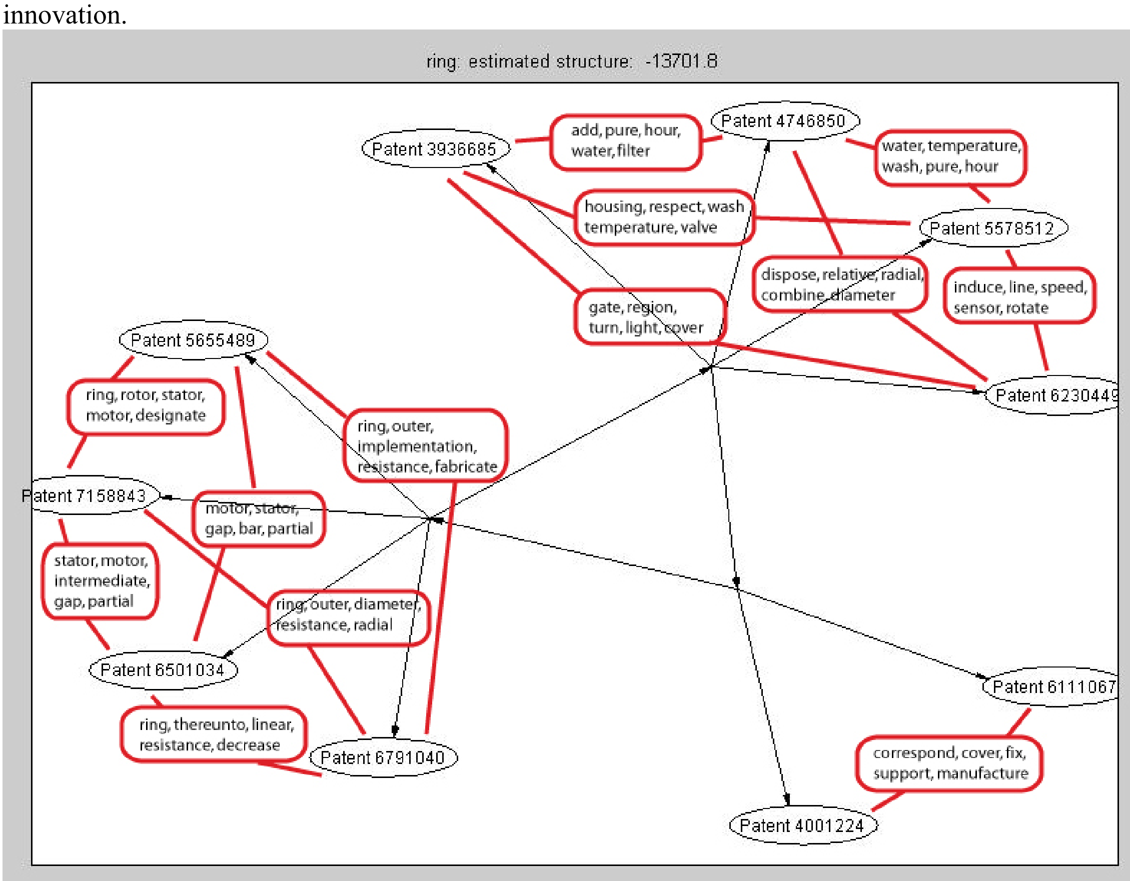

Figure 2. Ring Structure of Ten Random Patents Labeled with Five Most Similar Words Common to Patent Pair Connection as Derived from LSA

At a higher level, it is important to ground these structures in their potential and utility as design stimulation tools. Consider an example for how a designer might use these initial results: Suppose a designer is attempting to create a new product that will involve a motorized component or assembly. By examining the structure and finding the patents that involve similar parts or functionality, the designer could derive inspiration from different motor designs and methods of integration into larger assemblies or systems as found in these patents. This can be taken further than simple surface similarities by exploring patents within the structure that have labels indicating a desired functionality. This work will be expanded to find structures that are more intentionally feature based or function based by taking advantage of 'part of speech' parsing of the patents, which will hopefully lead to even richer results.

While it is interesting and meaningful to examine the "best" structural form to describe the data, the other forms are of interest as well. The other forms explored could lead to new ways of looking at the patents, specifically when comparing a more simplistic structure like a chain to a more complex structure like a tree. The more simple structure might distill the similarities of the patents to the absolute strongest, while the tree might allow for a much more complex or nuanced mapping. Exploring the meaning of the multiple structural forms that will be fit to the data is an important next step in understanding how this method will provide valuable output and potential inspiration to designers. In addition, depending on the input, the best structural form could be different for each data set. The goal of this work is not to find one "true" structure, but to uncover meaningful insights from the structures that emerge.

The value of the work presented here comes not only from the structures themselves, but from the combination of the structure finding algorithm with latent semantic analysis, yielding an automated method for facilitating the understanding, interpretation, and use of these structures to aid and inspire engineering designers.

5 CONCLUDING REMARKS

By applying an algorithm for discovering structural form to the U.S. patent database, this research intends to leverage this large repository of existing design solutions to generate analogical inspiration to engineering designers. The first steps toward that larger goal were presented here, through the application of a methodology not only to discover the structural form of a small set of random patents, but also to present the structures in a useful, efficient, and visually informative way using LSA. The first results from the presented methodology show meaningful results and promise insightful analysis of future, more complex structures and patent spaces.

REFERENCES

[1] Ullman, D. G., 2003, *The Mechanical Design Process, 3rd Edition,* McGraw-Hill Companies, New York, 158-159.

[2] "Cross-Industry Innovation," Business Process Optimization, European CEO, www.zuehlke.com/fileadmin/pdf/fachartikel/124_cd_european_ceo_cross_innovation.pdf, Accessed 9 April 2009.

[3] Basantani, M., "The MagLev: The Super-powered Magnetic Wind Turbine," Inhabitat Energy, http://www.inhabitat.com/2007/11/26/super-powered-magnetic-wind-turbine-maglev/, Accessed 9 April 2009.

[4] Linder, P., "Leveraging Novel MEMS Technologies for Next Generation Photovoltaic Applications," Semiconductor International, http://www.semiconductor.net/article/CA6576216.html, Accessed 11 April 2009.

[5] Herstatt, C. and Kalogerakis, K., "How to Use Analogies for Breakthrough Innovations," http://biblioteca.universia.net/html_bura/ficha/params/id/39800023.html, Accessed 11 April 2009.

[6] Rantanen, K., and Domb, E., Simplified TRIZ: New Problem Solving Applications for Engineers and Manufacturing Professionals, CRC Press, 2002.

[7] Griffiths, T. L., Kemp, C., and Tenenbaum, J. B., 2008, "Bayesian models of cognition" In R. Sun, editor, The Cambridge handbook of computational cognitive modeling. Cambridge University Press.

[8] Kemp, C. and Tenenbaum, J., "The Discovery of Structural Form," PNAS, 2008, Vol 105, no 31, 10687-10692.

[9] Dahl, D. W., and Moreao, P., 2002, “The Influence and Value of Analogical Thinking During New Product Ideation,” J. of Marketing Res., 39, February, 47-60.
[10] Goldschmidt, G., and Smolkov, M., 2006, “Variances in the Impact of Visual Stimuli on Design Problem Solving Performance,” Design Studies, 27, 549-569.
[11] Linsey, J. S., Wood, K. L., and Markman, A. B., 2008, “Modality and Representation in Analogy,” Artif. Intell. for Eng. Des., Anal. & Mfg., 22, 85-100.
[12] Christensen, B. T., and Schunn, C. D., 2005, “Spontaneous Access and Analogical Incubation Effects,” Creativ. Res. J., 17(2-3), 207-220.
[13] 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. & Cog., January, 35, 29-38.
[14] 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”, Design Studies, 29, 203-221.
[15] Moss, J., Kotovsky, K., and Cagan, J., 2007, “The Influence of Open Goals on the Acquisition of Problem-Relevant Information,” J. of Exp. Psych., 33(5), 876-891.
[16] Moss, J., Cagan, J., and Kotovsky, K., 2007, “Design Ideas and Impasses: The Role of Open Goals,” ICED (114), Paris, France, 1-10.
[17] Jansson, D. G., and Smith, S. M., 1991, “Design Fixation,” Design Studies, 12, 3-11.
[18] Smith, S. M., and Blankenship, S. E., 1991 “Incubation and the Persistence of Fixation in Problem Solving,” Amer. J. of Psych., 104 (1), 61-87.
[19] Purcell, A. T., and Gero, J. S., 1996, “Design and Other Types of Fixation,” Design Studies, 17, 363-383.
[20] Chrysikou, E. G., and Weisberg, R. W., 2005, “Following the Wrong Footsteps: Fixation Effects of Pictorial Examples in a Design Problem Solving Task,” J. of Exp. Psych.: Learn., Mem., & Cog., 31 (5), 1134-1148.
[21] Marsh, R. L., Ward, T. B. and Landau, J. D., 1999, “The Inadvertent Use of Prior Knowledge in a Generative Cognitive Task,” *Memory & Cognition*, 27 (1), 94-105.
[22] Smith, S. M., Ward, T. B. and Schumacher, J. S., 1993, “Constraining Effects of Examples in a Creative Generation Task,” *Memory & Cognition*, 21(6), 837-845.
[23] Stone, R. and Wood, K. L., 2000, “Development of a Functional Basis for Design,” J. of Mech. Des., Vol 122, 359 – 370.
[24] Cheong, H., Shu, L. H., Stone, R., and Wood, K. L., 2008, “Translating Terms of the Functional Basis into Biologically Meaningful Keywords,” Proceedings of the ASME International Design Engineering Technical Conference, New York, NY, USA, 3-6 August.
[25] Zhang, R., Cha, J., and Lu, Y., 2007, “A Conceptual Design Model Using Axiomatic Design, Functional Basis and TRIZ,” Proceedings of the 2007 IEEE IEEM, 1-4244-1529-2/07.
[26] Chakrabarti, A. K., Dror, I., and Nopphdol, E., 1993, “Interorganizational Transfer of Knowledge: An analysis of Patent Citations of a Defense Firm,” IEEE Transactions on Engineering Management, Vol. 40, No. 1.
[27] Indukuri, K. V., Ambekar, A. A., and Sureka, A., 2007, “Similarity Analysis of Patent Claims Using Natural Language Processing Techniques,” International Conference on Computational Intelligence and Multimedia Applications, 0-7695-3050-8/07, 13-15 December.
[28] Kasravi, C., and Risov, M., 2007, “Patent Mining – Discovery of Business Value from Patent Repositories,” Proceedings of the 40th Hawaii International Conference on System Sciences, 1530-1605/07.
[29] 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, 1094-7176/00, 48-55.
[30] Mukherjea, S., Bhuvan, B., and Kankar, P., 2005, “Information Retrieval and Knowledge Discovery Utilizing a BioMedical Patent Semantic Web,” IEEE Transactions on Knowledge and Data Engineering, Vol. 17, No. 8., 1099-1110.
[31] Koch, S., Bosch, H., Giereth, M., and Ertl, T., 2009, “Iterative Integration of Visual Insights during Patent Search and Analysis,” IEEE Symposium on Visual Analytics Science and Technology, Atlantic City, NJ, USA, Oct. 11-16.

[32] Chakrabarti, S., Dom, B., Agrawal, R., and Raghavan, P., 1998, “Scalable Feature Selection, Classification and Signature Generation for Organizing Large Text Databases into Hierarchical Top Taxonomies,” The VLDB Journal, 7, 163-178.
[33] Nosofsky, R. M., 1986, “Attention, similarity, and the identification-categorization relationship,” Journal of Experimental Psychology: General, 115:39–57.
[34] Ashby, F. G., and Maddox, W. T., 1992, “Complex decision rules in categorization: Contrasting novice and experienced performance,” Journal of Experimental Psychology: Human Perception and Performance, 18(1):50–71.
[35] Anderson, J. R., 1991, “The adaptive nature of human categorization,” Psychological Review, 98(3):409–429.
[36] Kemp, C., Tenenbaum, J. B., Griffiths, T. L., Yamada, T., and Ueda, N., 2006, “Learning systems of concepts with an infinite relational model,” In Proceedings of the 21st National Conference on Artificial Intelligence.
[37] Kemp, C., and Tenenbaum, J. B., “Structured statistical models of inductive reasoning,” Psychological Review, In press.
[38] Griffiths, T. L., Steyvers, M. , and Tenenbaum, J. B., 2007, “Topics in semantic representation,” Psychological Review, 114(2):211–244.
[39] Landauer, T.K., Foltz, P.W. & Laham, D., 1998, “Introduction to Latent Semantic Analysis,” Discourse Processes, 25, 259-284.
[40] Foltz, P.W., Kintsch, W., & Landauer, T.K., 1998, “The Measurement of Textual Coherence with Latent Semantic Analysis,” Discourse Processes, 25 (2&3), 285-307.
[41] Deerwester, S., Dumais, S. T., Furnas, G. W., and Landauer, T. K., 1990, “Indexing by Latent Semantic Analysis,” J. of Amer. Soc. of Info. Sci. & Tech., 41 (6), 391-407.
[42] Jaynes, E. T., 2003, Probability theory: The logic of science, Cambridge University Press, Cambridge, Chapter 1: Plausible Reasoning.
[43] Kemp, C. and Tenenbaum, J., “The Discovery of Structural Form,” PNAS, 2008, Supporting Information Appendix 10.1073/pnas.0802631105.