ICED13
INTERNATIONAL CONFERENCE ON ENGINEERING DESIGN, ICED13
19-22 AUGUST 2013, SUNGKYUNKWAN UNIVERSITY, SEOUL, KOREA

USING DESIGN DATABASE STRUCTURES TO CHARACTERIZE FREEDOM-TO-OPERATE IN A DESIGN SPACE: A LEGAL CASE STUDY

Katherine FU (1), James DILMORE (2), Jonathan CAGAN (3), Charles H. DOUGHERTY JR. (4)
1: MIT, United States of America; 2: Reed Smith LLP, United States of America; 3: Carnegie Mellon University, United States of America; 4: Beck & Thomas, P.C., United States of America

ABSTRACT

Novelty, and specifically freedom-to-operate (FTO), assessment is crucial step in launching and patenting a new product. We compare a traditional FTO analysis performed by a patent lawyer and expert in chemistry and pharmaceutical technology with the outcomes of a computational method. The computational method discovers the structural form of a set of patents using a text-derived similarity data, creating a descriptive space of the potentially relevant prior art. A test formulation of a fabricated new pharmaceutical drug was developed for the FTO and computational analysis, and strengths and areas for improvement of the computational method were identified. FTO analysis is time consuming and labor intensive, and results indicated that with further development, the computational method could aid patent lawyers in getting a faster and fresh snapshot of the space of prior art, and even point to patents most relevant to a proposed new product. Areas for improvement are intrinsic knowledge of the computational method in the field of application, and finding which sections of patents lead to most accurate representations of the space, and further automation and efficiency.

Keywords: novelty, patents, freedom-to-operate

Contact:
Dr. Katherine Kai-Se Fu
MIT
Mechanical Engineering
Cambridge
02139
United States of America
katherine.fu@gmail.com

1 INTRODUCTION

Novelty, defined as the quality of being new, original, or unusual (Merriam-Webster, 2013), is an attribute that is coveted in design and central to innovation. In the world of patents, an alleged invention must be novel to qualify for the exclusive rights afforded by patent protection. Separate and apart from whether an invention is patentable, prior to launching a new commercial product, a patent attorney will often conduct a freedom-to-operate (FTO) analysis to determine if the new product would infringe upon the exclusive patent rights of others (35 U.S.C. § 271). In this work, we examine one aspect of the current method of analyzing for freedom-to-operate in the context of current patent law practice and compare it to a proposed method for aiding in novelty evaluation using a computationally generated structure of the relevant patent space, with the ultimate aim of honing the computational method to increase efficiency in patent law practice. The computational method (Fu, 2012a) relies on three main steps: (1) preprocessing the textual content of the patent dataset to extract and quantify their contextual similarity to one another, (2) discovering the structural form of the patent similarity data using Kemp and Tenenbaum's algorithm (Kemp and Tenenbaum, 2008a, Kemp and Tenenbaum, 2008b), and (3) annotating the structure with labels for clusters and regions of the 2D patent space, connoting the meaning and logic behind the connectivity and layout. As we juxtapose the output of this computational method with the manual freedom-to-operate analysis, we attempt to strengthen the foundation for the use of the computational method as an aid to help patent lawyers be more resource efficient in their intellectual property work.

2 BACKGROUND

The following sections detail relevant literature regarding approaches to evaluating novelty in design research, and current methods for automated novelty detection using computational aids. Some background on the requirements and necessary procedures for determining freedom-to-operate, or novelty in patent law, is provided, as well as details of the technical background of the case study used for this analysis and comparison of the proposed computational structuring method, to the traditional approach.

2.1 Novelty Assessment in Design Research

Novelty is defined as the quality of being new, original, or unusual. A quality that is sought after in innovation and design, in the world of patents, it is essential to patentability and related to determining Freedom-to-Operate. Design methods researchers have created a number of different ways to evaluate novelty of design outcomes in order to test the effectiveness of ideation using new or trial design methods. Shah et al. (Shah, 2003) created a metric for novelty that relies on characterizing the design space with key attributes, and calculating novelty by using the inverse of the popularity of an instance of a particular solution method for each attribute. An alternative method to Shah's novelty metric is Amabile's (Amabile, 1982) Consensual Assessment Technique (CAT), which depends upon the agreement and judgment of observers with appropriate domain knowledge regarding the novelty or overall creativity of a design outcome. Another method for evaluating novelty, the Creative Product Semantic Scale (CPSS), has observers use a 7-point scale, evaluating three categories within novelty (original, surprising, germinal) with 5 subscales within each of the three categories (i.e. original -> conventional) (O'Quin and Besemer, 1989). Other methods include Comparative Creativity Assessment (CCA) (Oman et al., 2012), those developed by Sarkar and Chakrabarti (Sarkar and Chakrabarti, 2008), Evaluation of Innovative Potential (EPI) , and refinements on Shah's metrics by Nelson et al. (Nelson et al., 2009), Chulvi et al. (Chulvi et al., 2011), and Peeters et al. (Peeters et al., 2010), among even more. All of these methods involve human judgment, and while widely used, may be cumbersome and inefficient for use outside of ideation research, for example in applications like patent law freedom-to-operate assessments and searches of prior art in general.

2.2 Automated Novelty Detection

Automated novelty detection has been addressed by a number of researchers across many fields (Markou and Singh, 2003b, Markou and Singh, 2003a). Often, novelty detection takes the form of anomaly identification, for example in detecting masses in mammograms (Tarassenko, 1995) or radar target detection (Tax and Duin, 1998). Many fully automated approaches have been devised for novelty detection, and can be categorized into two main types, statistical approaches and neural network approaches (Markou and Singh, 2003b, Markou and Singh, 2003a). Most statistical approaches rely on modeling distributions of data sets, and estimating the probability of test data belonging to those distributions (Baker et al., 1999, Odin and Addison, 2000, Manson et al., 2000, Webb, 1999, Hellman, 1970, Fumera et al., 2000, Roberts and Tarassenko, 1994, Nairac et al., 1999, Yeung and Ding, 2002, Ruotolo and Surace, 1997, Guttormsson et al., 1999, Pizzi et al., 2001, Yang et al., 1998). Neural network approaches are much more computationally expensive, harder to retrain on new data, and can be more difficult to generalize (Markou and Singh, 2003b). Some key examples of neural network approaches to automated novelty detection include (Zhang and Veenker, 1991, Kwok and Yeung, 1999, Singh and Markou, 2004, LeCun et al., 1990, Augusteijn and Folkert, 2002, Tax and Duin, 1999, Diehl and Hampshire II, 2002, Granger et al., 1999, Brotherton and Johnson, 2001, Song et al., 2001, Kohonen, 1988, Kohonen, 2001, Aeyels, 1991, Barnett and Lewis, 1994, Bishop, 1994, Emamian et al., 2000, Marsland et al., 2000, Martinez, 1998, Tarassenko, 1995). The work presented in this paper approaches assessing novelty in a less automated, but more interactive and exploratory way. Our approach uses the visual structuring of data to aid an expert in efficiently examining the scope of the data, while still placing the final novelty evaluation in the hands of the expert. Other efforts have been made to visually organize data and/or patents for exploration and design practice, which include (Koch et al., 2009, Mukherjea et al., 2005, Chakrabarti et al., 1998, Verhaegen et al., 2011, CREAX, Goldfire, Forbus et al., 1994, Moehrle and Geritz, 2004, Souili and Cavallucci, 2012, Souili et al., 2011). These efforts have diverse computational bases, distinct from our efforts, and have yet to be used in the manner we propose with this work.

2.3 Novelty and Freedom to Operate in Patent Law

Novelty and freedom to operate are two related, but distinct, concepts within U.S. patent law. The patentability of an invention is established when it satisfies the three-fold statutory requirements of being directed to patentable subject material (35 U.S.C. § 101), and being both novel (35 U.S.C. § 102) and non-obvious (35 U.S.C. § 103) over the prior art – i.e. that which has come before the current invention. A patent grants its holder the right to exclude others from making, using, selling, offering to sell, or importing into the United States the claimed invention (35 U.S.C. § 271(a)). The scope of the patent holder's exclusionary right is defined by the claims of an issued patent.

As noted above, a patent grants the patent holder the right to <u>exclude</u> others from undertaking certain activities (i.e. making, using, selling, offering for sale or importing the invention). A patent does not, however, grant the patent holder the right to <u>practice</u> (make, use or sell) the claimed invention as third parties may have patent coverage that dominates the patentee's position. An invention may be patentable over the prior art (i.e., novel and non-obvious), but still include elements or components that are themselves patented. Thus, prior to commercializing a new product or undertaking a new method of manufacture, companies will often engage a law firm to conduct an assessment of whether it can undertake those commercial activities without infringing an issued patent – that is, whether the company has "freedom to operate." The initial step in such an evaluation is to conduct a search of the prior art to identify those patents most closely related to the commercial embodiment. The attorney then determines whether the proposed commercial embodiment would infringe the claims of any of the patents through comparing the claim scope to the proposed product or method.

2.4 Case Study Background

To evaluate the ability of computational methods to contribute to the process of conducting a freedom-to-operate analysis, we selected a commercially relevant corpus of patents on which a test case study could be performed. The pharmaceutical company Purdue has developed a diverse and extensive patent portfolio relating to its pharmaceutical products. We selected a portion of Purdue's portfolio on which to conduct a case study without regards to expiration date of the patent, patent term extension, or other complicating factors. The test "commercial embodiment" was sustained-release Oxycontin® - a successful pharmaceutical formulation that is presently commercially available from Purdue. While this arrangement is not typical of a freedom-to-operate analysis, we chose Purdue's own product so as not to implicate any competitors or potential future products in this highly litigious area.

The patent corpus was selected so as to include some patents that were very relevant to Oxycontin® and others that were entirely unrelated to that formulation. Oxycontin® contains an opioid agonist (i.e., oxycodone) in a solid tablet formulation that is to be ingested orally. The formulation is to be taken twice daily to achieve sustained analgesia in patients. Purdue's patent portfolio includes many patents that encompass such formulations and others that are closely related (e.g., sustained-release formulations where a different opioid agonist, such as morphine, is employed). Other, marginally related patents include ones covering abuse-deterrent opioid oral formulations and transdermal opioid formulations. Because of a wide-ranging research program, Purdue also holds patents on unrelated pharmaceutical formulations such as non-opioid sustained release compositions (e.g., for the stimulant methylphenidate) and injectable formulations. The corpus for the test case here encompassed that breadth of technical detail to provide both "close calls," where the computational analysis would need to finely distinguish between similar patents and "easy calls," where the claimed formulations or methods were obviously distinct from the target commercial embodiment of sustained-release Oxycontin®.

3 METHODS

The results of two methods for understanding novelty within a patent space are presented in this paper. This section details the steps taken to produce the results of the two methods – the manual method for freedom-to-operate analysis and the computational method for generating structures of design databases.

3.1 Manual Method: Freedom-to-Operate Analysis

In the present example there are many approaches that an attorney could take to narrow the initial comprehensive collection of patents to the most relevant ones that require additional review. In all instances, the initial review would focus on the patent claims – the numbered sentences that appear at the very end of an issued patent and define the scope of patent protection. During the initial pass, the attorney would identify at least one claim element that is clearly not present in the proposed commercial product (i.e., the product that is the focus of the Freedom-to-Operate (FTO) analysis). For example, if the proposed product was a pharmaceutical formulation for oral administration, then all patents from the comprehensive collection of patents that were directed to injectable or transdermal administration could be eliminated immediately. Similarly, if the active ingredient of the proposed product was delivered in a sustained-release formulation, all patents limited to immediate-release formulations could likewise be eliminated. This process continues until a limited group of most relevant patents are identified. Those patents are considered in greater detail by carefully evaluating the meaning of claim terms in light of the patent description and the prosecution history. During a full legal analysis, fine differentiating points between the relevant patent claims and the commercial product may rest on nuanced interpretations of claims based on technical expertise (e.g., chemistry, biology) or legal principles (e.g., estoppel, claim construction). The limitations of this approach stem from the large amount of time and effort (and thus, expense) of the undertaking.

3.2 Computational Approach: Design Database Structure Generation

There are three parts of the computational method for producing the structures presented in the results section of this paper. First, similarity data characterizing a space of patents is created using Latent Semantic Analysis (LSA). Second, an algorithm developed by Kemp and Tenenbaum (Kemp and Tenenbaum, 2008b) is employed to then discover the structural form(s) of the patent data using Bayesian inference, with the output from LSA as input. Finally, to characterize and understand the cluster of patents within the structures produced in the second step, LSA is again used to create labels for the clusters of patents, based on word to document similarity calculations. More detail on this method can be found in (Fu, 2012a, Fu, 2012b).

Step 1: Pre-processing with LSA

Similarity data for characterizing a space of patents is generated using LSA, which is then input into the structural form discovery algorithm. Briefly, LSA is a computational text analysis tool that can extract the contextual similarity of documents and/or words (Landauer et al., 1998, Foltz et al., 1998, Deerwester et al., 1990). It involves four main steps: 1. Creation of a word-by-document matrix; 2. An inverse "entropy weighting" step, giving 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; 4. Calculation of the cosine similarity between documents, yielding a matrix of document-to-document coherence values. (Landauer et al., 1998, Foltz et al., 1998, Deerwester et al., 1990). The 87 patents in the current analysis were chosen by the second author as a relevant set within which to explore novelty and freedom-to-operate, all within the portfolio of Purdue. Patent families were generated by identifying patents that shared a common priority document. Given the set of 87 patents, either the full text or only the claims section of the patents is first parsed from HTM 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 output of the LSA step of the methods was two separate similarity matrices, one using the full text of the patents and one using only the claims of the patents. After applying the second and third step in the method, the full text similarity data led to the results in Figures 1 and 2, while the claims data led to the results in Figures 3 and 4.

Step 2: Discovery of Structural Form

Kemp and Tenenbaum's algorithm for discovering structural form in data (Kemp and Tenenbaum, 2008b) uses Bayesian inference to find natural structure in sets of data, evaluating how well each of the eight candidate forms describes the input inter-relationships among the entities within the data. The candidate forms include the partition, chain, order, ring, tree, hierarchy, grid, and cylinder. The structures have the potential to indicate underlying relationships among the patent data, as they have done with many other example data sets as presented in Kemp and Tenenbaum's work (Kemp and Tenenbaum, 2008a, Kemp and Tenenbaum, 2008b), making them apt to uncover information regarding freedom-to-operate in the space. For brevity, we will not dive deeply into the algorithm. The result of the Kemp and Tenenbaum algorithm is the best structure (instantiation) of each candidate form of the input patent data, and the associated posterior probability. Using the posterior probability values, the best structure can be identified. In the results section, we focus on the best structure as determined by the algorithm, a tree.

Step 3: Cluster Labeling with LSA

The third step in the method again employs LSA. This step is intended to create an automated way to enable analyzing the meaning of the connections between patents in the structures with a characterization of the clustering of patents. Latent Semantic Analysis is used to find the words in the LSA space that had the highest cosine similarity value to each patent. The output from this procedure was then used in two different ways. 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 abstract set of functionalities or terms. 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 not specific enough information to be useful. The second method, the Highest Cosine Similarity labeling method, generates labels 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 in the cluster, 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 clustering. The two methods are juxtaposed in the structures presented in the results, providing both sets of benefits, while mitigating the drawbacks of one another simultaneously. Improvement of these labeling methods is an area of future research.

Test Formulation Placement

A test formulation, here a description of a possible new drug, was added to the LSA space in addition to the 87 patents, the text of which was the following:

The test formulation contains oxycodone as the opiate agonist. It is an extended release formulation that allows administration twice a day, i.e. every twelve hours. The extended release is achieved by a dual-polymer matrix that is used within the formulation. The formulation also includes: ammonio methacrylate copolymer, hypromellose, lactose, magnesium stearate, polyethylene glycol 400, povidone, sodium hydroxide, sorbic acid, stearyl alcohol, talc, titanium dioxide, and triacetin as inactive ingredients. Agonist: oxycodone; Formulation type: oral extended release; Frequency of administration: twice daily; Extended release mechanism: Matrix and achieved by the ammonio methacrylate copolymer.

The purpose of this step was to locate what can be thought of as a "starting point" in the space, identifying the patents in the space that are most similar to the test formulation by calculating the average cosine similarity of the set of patents in each cluster to the test formulation, and selecting the cluster with the highest average cosine similarity to the test formulation as the starting point. This cluster node is marked with a grey filled circle in the figures shown in Section 4.1.

Region Identification
The second author determined the boundaries and labels for the regions contained in structures shown in Figures 1 and 2, marked with red lines and red text. These regions are based on expertise in the field of application of the case study and an advanced knowledge of chemistry and pharmaceutical formulations, and were determined by examining each patent in detail, and evaluating the scope of the patent disclosure or claims, as appropriate. In future work, the authors will seek an automated way of generating region boundaries and labels.

4 RESULTS

The results presented juxtapose the computational findings with the manual freedom-to-operate analysis, examining the utility of the computational method in relation to current patent law practices, and the potential benefits, flaws, and areas for improvement.

4.1 Computational Findings

Figures 1 and 2 show the output of the computational method presented in Section 3.2 using the full text of the patents in Figure 1, and the claims text only in Figure 2. The FTO analysis was performed on the full set of 106 patents. However, due to extremely high similarity of the full text of the patent to another patent in the set, as determined by Latent Semantic Analysis document cosine similarity, only 87 of these patents were used in the computational structuring method. Figures 1 and 2 show the structural form type deemed "best" based on its posterior probability score in both the full text and claims only analysis, which was the tree. The test formulation was placed in each figure at the node with the gray circle behind it. Dark gray numbered circles are the patents which were identified as most relevant to the test formulation based on the FTO analysis, while light gray numbered circles are the patents identified as partially or potentially relevant to the test formulation from the FTO analysis, as discussed below. One yellow circle identifies a patent that was not relevant to the test formulation at all, but accidentally included due to a typo in the patent number, 15(3). The red lines and text denote manually identified regions, as described in Section 3.2. When evaluating the full text of the patents, the algorithm tended to group patents based on major salient features of the disclosure, as expected. For example, Figure 1 shows that the algorithm identified large groups of patents that were directed to abuse-deterrent formulations of opiates (large dotted region in bottom left quadrant) and for formulations of opiates that are based opiate-coated small beads or particles (large dotted region in bottom right quadrant). Interestingly, the algorithm reached across patent families to identify numerous, unrelated patents that possessed conceptually related disclosures (e.g., patents from families 1, 8, 9, 10, 63, 69, and 70 in the small bead-containing formulation group). In contrast, the groupings were not as tight or readily apparent when the algorithm evaluated the patent claim text only as seen in Figure 2. That result is to be expected given the smaller content of claims when compared to the entire patent disclosure. The authors also suggest that the greater fragmentation of the patent claim space was due to more distinct differentiating points in the claims (which are carefully crafted to capture subtle points of patentability) when compared to the entire specification (the scope of which is significantly broader when compared to patent claims).

4.2 Manual Method: FTO Findings

The FTO analysis was performed on a set of 106 patents. The second author evaluated the claims of each of the 106 patents, first, in an effort to discard patents as irrelevant to the proposed commercial embodiment. For example, patents 62(2) and 62(3) were disregarded because the active component of the formulation was methylphenidate, which is not present in the proposed commercial formulation. Similarly, patents 67(1) (diamorphine), 72(1) & 72(2) (hydromorphone), and 4(1) & 4(2) (morphine) were considered irrelevant as claiming a different opiate agonist. Other patents were rejected as containing elements not present in the proposed formulation (e.g., 5(1)-5(5); opioid antagonist), having a specially treated coating (e.g., 1(3), 1(8), 1(11); plasticized ethylcellulose), or being in a different formulation (e.g., 12(1)-12(3); transdermal buprenorphine). After undertaking that process under numerous dimensions, five patents were identified as relevant to the test formulation: 14(1), 14(2), 14(3), 80(1), and 81(1). Due to extremely high similarity of the full text of the patent to another

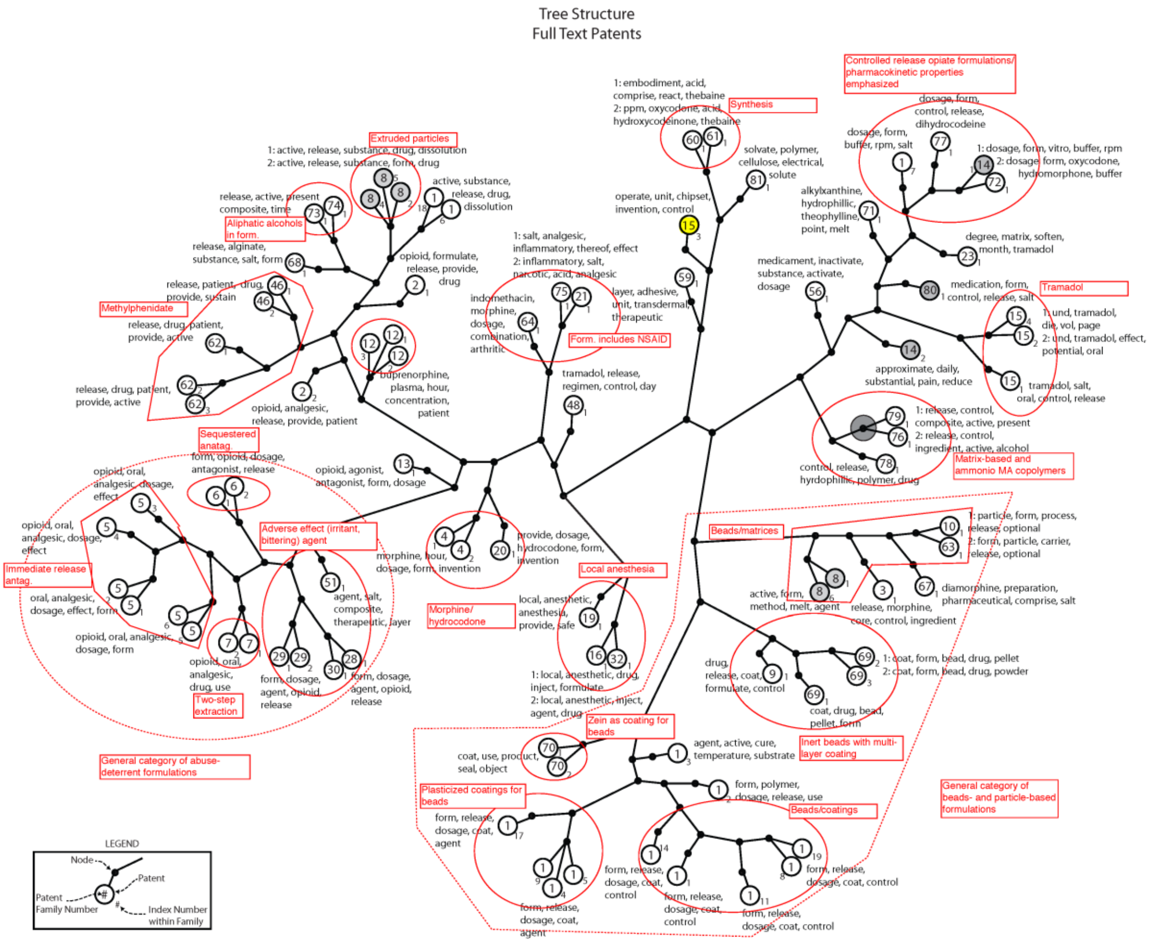

Figure 1. Tree Structure using Full Text of Patents

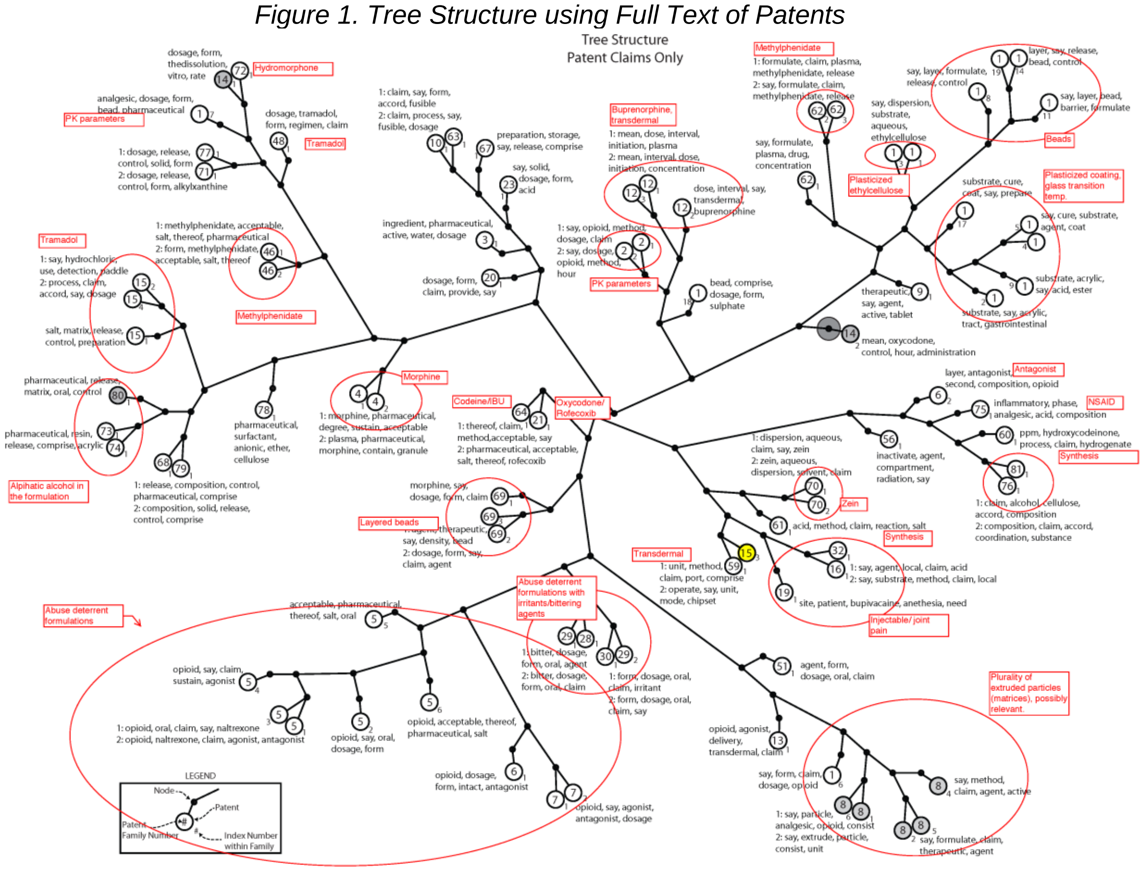

Figure 2. Tree Structure using Only Claims Section of Patents

patent in the set, only three of these five were included in the computational analysis and can be seen in the structures: 14(1), 14(2), and 80(1), which are indicated as darker gray filled numbered circles in the figures. These patents were identified as most relevant because their claims were drawn to oral tablet formulations that either explicitly or generically (e.g., "pharmaceutically active agent"; 80(1)) recited oxycodone as agonist and had the specified components of the commercial embodiments.

A secondarily relevant set of seven patents was identified in the FTO analysis, which might be relevant or are on the cusp of relevance: 1(12), 1(13), 8(1),8(2), 8(4), 8(5), 8(6). Again, only five of these were included in the computational analysis for the same reasons: 8(1), 8(2), 8(4), 8(5), 8(6). These are indicated in each of the structures as lighter gray filled numbered circles. These patents were identified as potentially relevant because their claim scope did not exclude the proposed commercial embodiment, and the claims included limitations that may or may not be possessed by the proposed formulation. Further assessment would be required to provide a definitive answer on whether the patents were relevant.

4.3 Analysis and Comparison of Methods

The comparison of the computationally generated structures to the FTO findings was both validating in parts, and indicating of areas for improvement and added complexity to the computational method. In validation, there are subtly identified subgroups that have been correctly clustered by the computational method; for example, in Figure 2, the branch toward the bottom left including families 5, 6, and 7 are all patents dealing with abuse-deterrent formulations, and the branch including families 1 and 8 are all patents relating to bead-based formulations. In examining the full text structures as compared to the claims only structures, two trends emerged. We noticed that, with the full text data, the clustering of the patents emerged much more cleanly and based on the semantic content of the patents in addition to the pre-existing structure within the patent database. For example, the branch encompassing the bead-based formulations obtained using the full text data contained a larger number of patents from diverse families when compared to the claims-only data set. Additionally, there were more "stragglers" (ungrouped patents) in the claims-only data set. We suggest this is due to the nature of claims (which are drafted to capture subtle differences) and the reduced size of the corpus, when compared to the full text of patents.

By effectively collecting patents across numerous families that were semantically related, the algorithm is able to aid in a legal FTO analysis by identifying patents that are <u>not</u> relevant and not worthy of further analysis. Conversely, when evaluating patent claims, the algorithm was successful in identifying the most relevant patents. During an FTO analysis, an attorney could first evaluate those patents identified by the algorithm as most relevant. Those patents are likely to be of high value in conducting an FTO analysis.

Some insights were gained as to key areas for expansion and improvement of the computational method. It was evident that there is a level of background knowledge within the field of chemistry that is implicit in the patents that were examined – principles that cannot be extracted by looking at the textual content of the patents, but an expert in chemistry would already know by simply looking at the patents. For example, the term "aliphatic alcohol" encompasses a class of compounds that possess specific chemical characteristics. Each member of that class may not be recited in the patent, but one of skill in the chemical arts would recognize the members of that chemical class. Similarly, chemistry abounds with numerous synonyms for a single compound, complicating interpretation of the plain text. For example, the chemical compound "hydroxypropyl methyl cellulose" is also known as "hypromellose." Such information may be entirely absent from a patent application, but one of skill in the chemical arts would be familiar with such nomenclature. We see this as a clear opportunity to add a further context to the space, to make connections and clustering richer, better informed, and closer to expert thinking and understanding of the space. This could be achieved through adding seminal chemistry background texts to the corpus when performing LSA, while leaving them out of the actual structuring step, to build context.

When conducting an FTO analysis, the initial challenge is starting with a comprehensive collection of patents (obtained from an expansive search of the prior art) and appropriately narrowing that comprehensive list to a subset of patents that require additional review by a knowledgeable patent attorney. Computational methods like the one described herein provide one approach to undertaking this challenge in an efficient manner. For example, after first characterizing the entirety of the comprehensive collection it would be routine to group and re-group all the patents around particular attributes of a proposed commercial product to identify a subset of patents that are most relevant to that specific product. Conceptually, this methodology could be applied to an initial collection of patents of any size, limited only by the computational power available to the user. Another limiting factor affecting how well a given collection of patents could be characterized would depend upon the consistent use of terminology (as chosen by the patent drafters) in relationship to the "intelligence" of the system in recognizing synonymous wording and concepts. This limitation on "system intelligence" would likely diminish over time as the operator was able to introduce ("teach") synonyms and related concepts to the process.

5 CONCLUSIONS

Freedom-to-Operate in patent law is a necessary analysis in bringing a new product to market, but is a very time and energy intensive process, needing attention to nuance and detail by expert minds. We juxtapose a FTO analysis with a computational method to discover structure in a prior art space of patents, and examine the benefits and areas for improvement of the computational method based on the case study performed. The computational method was able to effectively cluster patents across numerous families that were semantically similar. It has potential aid in a legal FTO analysis by both identifying patents that are not relevant to a test formulation, and identifying patents that are most relevant to the test formulation. Some areas identified for future work are expanding the "expertise" of the computational method to pick up on more nuanced or latent characteristics of the patents and language used within them, as well as increasing the efficiency and automation of the method. This work contributes to the field of design science through applying computational techniques to aid in the efficacy of intellectual property work, a necessary but often arduous task in new product development. These same techniques can also be extended in a more general way to novelty assessment and even conceptualization in a design space and process.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation, under grant CMMI0855326.

REFERENCES

Amabile, T. M. 1982. Social Psychology Of Creativity: A Consensual Assessment Technique. *Journal Of Personality And Social Psychology,* 43, 997-1013.

Baker, L. D., Hofmann, T., Mccallum, A. K. & Yang, Y. 1999. A Hierarchical Probabilistic Model For Novelty Detection In Text. Technical Report.

Chakrabarti, S., Dom, B., Agrawal, R. & Raghavan, P. 1998. Scalable Feature Selection, Classification And Signature Generation For Organizing Large Text Databases Into Hierarchical Topic Taxonomies. *The Vldb Journal* 7, 163-178.

Chulvi, V., Mulet, E., Chakrabarti, A., López-Mesa, B. & González-Cruz, C. 2011. Comparison Of The Degree Of Creativity In The Design Outcomes Using Different Design Methods. *Journal Of Engineering Design,* 23, 241-269.

Creax: Creativity For Innovation [Online]. Available:Http://www.Creax.Com [Accessed 7 Sept. 2012.

Deerwester, S., Dumais, S. T., Furnas, G. W. & Landauer, T. K. 1990. Indexing By Latent Semantic Analysis. *Journal Of The American Society For Information Science* 41, 391–407.

Foltz, P. W., Kintsch, W. & Landauer, T. K. 1998. The Measurement Of Textual Coherence With Latent Semantic Analysis. *Discourse Processes* 25, 285-307.

Forbus, K. D., Gentner, D. & Law, K. 1994. Mac/Fac: A Model Of Similarity-Based Retrieval. *Cognitive Science*, 19, 141-205.

Fu, K. 2012a. *Discovering And Exploring Structure In Design Databases And Its Role In Stimulating Design By Analogy.* Ph.D. Dissertation, Carnegie Mellon University.

Fu, K., Cagan, J., Kotovsky, K., Wood, K. 2012b. Discovering Structure In Design Databases Through Function And Surface Based Mapping. *Asme Journal Of Mechanical Design,* In Press.

Goldfire, I. M. *Invention Machine Goldfire: Unleashing The Power Of Research* [Online]. Available: Http://Inventionmachine.Com/Products-And-Services/Innovation-Software/Goldfire-Research/ [Accessed 19 February 2012.

Kemp, C. & Tenenbaum, J. 2008a. The Discovery Of Structural Form. *Pnas*, Supporting Information Appendix 10.1073/Pnas.080263110.

Kemp, C. & Tenenbaum, J. B. 2008b. The Discovery Of Structural Form. *PNAS USA,* 105, 10687-92.

Koch, S., Bosch, H., Giereth, M. & 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.
Landauer, T. K., Foltz, P. W. & Laham, D. 1998. An Introduction To Latent Semantic Analysis. *Discourse Processes* 25, 259-284.
Markou, M. & Singh, S. 2003a. Novelty Detection: A Review—Part 1: Statistical Approaches. *Signal Processing,* 83, 2481-2497.
Markou, M. & Singh, S. 2003b. Novelty Detection: A Review—Part 2: Neural Network Based Approaches. *Signal Processing,* 8, 2499 – 2521.
Merriam-Webster 2013. Novelty. *Merriam-Webster*.
Mukherjea, S., Bhuvan, B. & Kankar, P. 2005. Information Retrieval And Knowledge Discovery Utilizing A Biomedical Patent Semantic Web. *Ieee Transactions On Knowledge And Data Engineering* 17, 1099-1110.
Nelson, B. A., Wilson, J. O., Rosen, D. & Yen, J. 2009. Refined Metrics For Measuring Ideation Effectiveness. *Design Studies,* 30, 737-743.
O'quin, K. & Besemer, S. P. 1989. The Development, Reliability, And Validity Of The Revised Creative Product Semantic Scale. *Creativity Research Journal,* 2, 267-278.
Oman, S. K., Tumer, I. Y., Wood, K. L. & Seepersad, C. 2012. A Comparison Of Creativity And Innovation Metrics And Sample Validation Through In-Class Design Projects. *Res. in Eng. Design*.
Peeters, J., Verhaegen, P.-A., Vandevenne, D. & Duflou, J. R. 2010. Refined Metrics For Measuring Novelty In Ideation. *Idmme - Virtual Concept.* Bordeaux, France.
Sarkar, P. & Chakrabarti, A. 2008. The Effect Of Representation Of Triggers On Design Outcomes. *Artif Intell Eng Des Anal Manuf,* 22, 101-116.
Shah, J. J., Vargas-Hernandez, N., Smith, S. M. 2003. Metrics For Measuring Ideation Effectiveness. *Design Studies,* 24, 111-134.
Tarassenko, L. 1995. Novelty Detection For The Identification Of Masses In Mammograms. *4th Iee International Conference On Artificial Neural Networks.*
Tax, D. M. J. & Duin, R. P. W. 1998. Outlier Detection Using Classifier Instability. *Advances In Pattern Recognition, The Joint Iapr International Workshops,* 593-601.
Verhaegen, P., D'hondt, J., Vandevenne, D., Dewulf, S. & Duflou, J. R. 2011. Identifying Candidates For Design-By-Analogy. *Computers In Industry,* 62, 446-459.