

DETC2014-34491

FACILITATING DESIGN-BY-ANALOGY: DEVELOPMENT OF A COMPLETE FUNCTIONAL VOCABULARY AND FUNCTIONAL VECTOR APPROACH TO ANALOGICAL SEARCH

Jeremy Murphy
Schlumberger Limited
Sugarland, TX, USA

Katherine Fu
Mass. Institute of Technology
Cambridge, MA, USA
& Singapore University of
Technology and Design
Republic of Singapore

Kevin Otto
Engineering Product Development Pillar
Singapore University of
Technology and Design
Republic of Singapore

Maria Yang
Mass. Institute of Technology
Cambridge, MA, USA

Dan Jensen
United States Air Force Academy
Colorado Springs, CO, USA

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

ABSTRACT

Design-by-analogy is an effective approach to innovative concept generation, but can be elusive at times due to the fact that few methods and tools exist to assist designers in systematically seeking and identifying analogies from general data sources, databases, or repositories, such as patent databases. A new method for extracting analogies from data sources has been developed to provide this capability. Building on past research, we utilize a functional vector space model to quantify analogous similarity between a design problem and the data source of potential analogies. We quantitatively evaluate the functional similarity between represented design problems and, in this case, patent descriptions of products. We develop a complete functional vocabulary to map the patent database to applicable functionally critical terms, using document parsing algorithms to reduce text descriptions of the data sources down to the key functions, and applying Zipf's law on word count order reduction to reduce the words within the documents. The reduction of a document (in this case a patent) into functional analogous words enables the matching to novel ideas that are functionally similar, which can be customized in various ways. This approach thereby provides relevant sources of design-by-analogy inspiration. Although our implementation of the technique focuses on functional descriptions of patents and the mapping of these functions to those of the design problem, resulting in a set of analogies, we believe that this technique is applicable to other analogy data sources as well. As a verification of the approach, an original design problem for an automated window washer illustrates the distance range of analogical solutions that can be extracted, extending from very near-field, literal solutions to far-field cross-domain analogies. Finally, a comparison with a current patent search tool is performed to draw a contrast to the status quo and evaluate the effectiveness of this work.

1 INTRODUCTION

Design-by-analogy using computational support methods offers a means to expand the set of considered concepts to entire online databases of concepts. The objective of this research is to develop appropriate algorithms and tools to enable web-based search for design analogies. With such an approach, a designer will be able to methodically search the vast amount of design information available online in patent archives. The resulting analogous concepts will be used to complement and infuse the concept generation process by introduction of non-obvious analogies resulting in innovative conceptual designs.

This research utilizes previous work encompassing functional modeling and representation of design concepts, online information retrieval from text-based databases, and concept similarity metrics to develop a systematic method for extracting near- and far-field analogies based on functional similarity [1-15]. Our hypothesis is that a patent-based analogy search algorithm utilizing a functional representation in a formalized tool can be used to identify non-obvious functional analogies for design concept generation more effectively than traditional key word searching for analogous designs. A key feature here is to make use of functional representations rather than component form or conflict representations. The underlying hypothesis is that functional analogies are useful within the conceptual design process to improve ideation outcomes.

1.1 Functional Modeling

Complementary to many design methodologies and philosophies, the use of functional analogies has been promoted as an important technique for synthesizing innovative and novel solutions, and research has empirically verified this effect [16, 17]. For effective application of these functional analogy based concept generation techniques, design problems may be represented by a set of solution-neutral functions to minimize design fixation and enable a large number of concepts to be considered [8, 15, 18].

Otto and Wood present a process of developing a functional representation of a concept beginning with an abstracted black box formulation of the overall product function with input and output flows

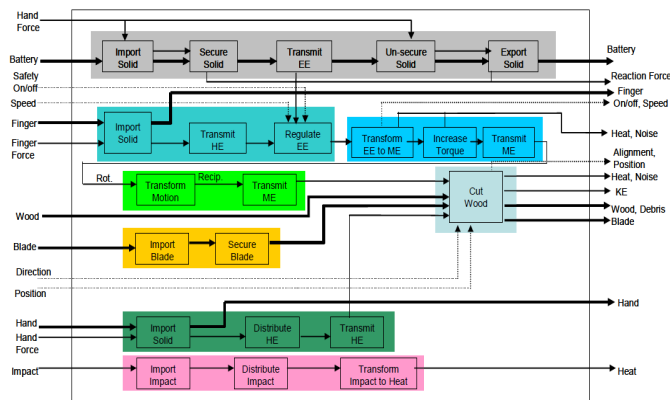


FIGURE 1: FUNCTIONAL MODEL FOR THE BLACK AND DECKER FIRESTORM JIGSAW.

[8]. The black box model is then decomposed into sub-functions interconnected by associated flows, resulting in a repeatable function structure representing the internal functionality of a concept [4, 5]. A number of functional models may be developed for a given design problem, depending on process choices and associated flows [8]. As an example, Figure 1 illustrates an exemplified functional model of a Jigsaw system. This example will be revisited in this paper.

Extensive efforts have produced a standard language for representing functions and flows associated with each sub-function called the *functional basis* [6, 7]. The functional basis consists of a set of function and flow words used as *verb-object* couples to describe the action imparted on the flows a function. Pahl and Beitz showed all function verbs can be abstracted hierarchically into more abstract function verbs, generally five overall “categorical functions” [15]. This set of functions forms a basis that can thereby provide a standard taxonomy for describing a design concept and enables physical systems, concepts or products to be functionally represented and compared.

Functional modeling is also used to identify modules and interface boundaries, such as that devised by Stone et al. to transform product function into alternative product layouts by identifying modules for modular product architectures [19]. This information can be used to simplify a complex functional model as well as discover opportunities to improve manufacturability, maintainability, and reliability early in the design process through function sharing and proper interface design, such as in [20]. The modules contained in the Jigsaw product are signified by the function chains contained in each colored box in Figure 1.

For our purposes here, a standardized functional model also facilitates archiving and retrieval of design knowledge. To that end, several systems have been developed to store the design knowledge contained in the functional models for design reuse [21-23]. In addition, computational tools have been developed to exploit the knowledge contained in the design repositories for the purpose of concept variant generation [24-26]. These works lend credence to the assertion that functional modeling is a useful engineering language for indexing design concepts.

One limitation of the functional modeling methodology of Otto and Wood [8] is that process choices made initially about what kind of inputs will go into the system on the user side necessitates selection of flow variables early in the conceptual design process [8]. In the Jigsaw

example, a process choice is made to utilize electricity and stored electrical energy (battery) for the power source. This is an obvious choice for a portable device, but alternative power sources such as fuel cells, solar cells, or pneumatically powered devices are expressly excluded from consideration. The single, domain-dependent model can lead to missed opportunities for novelty and innovation. Instead functional modeling should be considered more broadly in the context of modeling user and environmental activities and functions, and alternative process choices considered through higher levels of abstraction that lead to alternative functional models [8]. We explore this through matching using the functional basis approach.

1.2 Analogical Reasoning

Understanding the cognitive process involved in forming analogies is fundamental to the development of any tool or methodology that seeks to improve the conceptual design process. Analogy can be viewed as a mapping of knowledge from one situation (source) to another (target), enabled by a supporting system of relations or representations between situations [27-29]. The process of analogical comparison fosters new inferences and promotes construing problems in new insightful ways. The potential for creative problem solving is most noticeable when the situation domains are very different [16, 30]. The cognitive analogical process is based on the representation and processing of information, and therefore can be implemented in algorithms given an appropriate representation of the information [31, 32].

One method with great potential to produce innovative design is *design-by-analogy*. Previous research has shown usage of analogy can mitigate the effects of design fixation [33]. Theoretically, a robust design-by-analogy methodology would enable designers to identify non-obvious analogous solutions, even in cases where the mapping between concepts is tenuous or the concepts are from different domains. Such different but analogous concepts can be identified by creating abstracted functional models of concepts and comparing the similarities between their functionality. Appropriate functional representation of design concepts is as critical to the successful implementation of design-by-analogy as is developing a systematic approach to search for and evaluate the utility of functionally similar concepts.

Hacco and Shu also developed structured approaches utilizing biomimetic principles for generating concepts, which provides a systematic process for identifying analogous concepts [34]. They also use a functional semantic representation, in which keywords are derived that relate the function to the biological processes. A search is then performed using standard biological processes from biology textbooks as the reference database. Goel et al. also use functional modeling and functional indexing to create a system called KRITIK that autonomously generates new conceptual designs based on a case library of previously existing designs [35]. Bhatta et al. developed a project called IDeAL, which uses a function-behavior-structure model-based approach to design-by-analogy through pattern finding, constraint analysis, and problem reformulation [36]. Navinchandra et al. developed a nonfunctional approach using case based reasoning in a tool called CADET, to retrieve and synthesize case design components for more effective combination and better design [37]. Qian and Gero created an exploration medium for between-domain analogies using function-behavior-structure design prototypes [38]. Charlton and Wallace created a web-based tool for finding pre-existing engineering components for reuse in non-standard applications in new designs to reduce manufacturing costs [39]. FunSION, a computational tool developed by Liu et al., takes qualitative functional input and output requirements, and generates physical embodiments of design solutions [40, 41]. Chakrabarti et al. created Idea-Inspire, a database and software tool that automates analogical search in a natural and

artificial systems database to provide inspiration in the design process [42, 43]. Yang et al. worked to create thesauri using information retrieval from informal design documentation for reuse in the design process [44, 45], in addition to creating the DedalAI system to automatically index design concepts in electronic notebooks for retrieval and reuse [46]. Ahmed developed a system for helping designers to index and build a knowledge network based on engineering designer queries, which generates associations between concepts, with the end goal of aiding in the search for information, reformulation of a query, and prompting design tasks [47]. Linsey et al. [48-50], Seger et al. [51], and Verhaegen et al. [52] develop approaches to analogical retrieval and reasoning through linguistic associations, problem re-representation, and mappings.

Linguistics research has shown that verbs are inherently relational by nature and impose fewer psychological constraints compared to nouns [53]. Verbs represent relational concepts whereas nouns are object-reference concepts. In the following section, functional representation making use of verbs is proposed for design problems to leverage the cognitive flexibility of the action verb.

As discussed previously, the functional modeling approach of Otto and Wood using the function-flow basis is a useful method for representing design intent, but the specification of the flow inputs has the consequence of defining the solution domain due to the process choices. A truly solution-independent representation should not fix the design space within a particular domain. Besides the exploration of multiple, simultaneous functional models, this can be accomplished by removing the flow objects from the verb-object functional model, and focusing on the verbs. The resulting abstract verb-based representation is entirely conceptual, relational and solution independent. For example, the functional semantic representation of the Jigsaw in Figure 1 can be expressed as abstract basis functions:

Import : Secure : Release : Transmit : Regulate : Transform : Distribute

The representation is greatly simplified as common functions acting on different flows collapse into a single verb. It is acknowledged that the semantic representation as expressed in abstracted basis functions lacks the granularity necessary to be useful for concept generation [54]. The representation scheme will use lower level function (tertiary and correspondent) to specify the design problem. The Jigsaw translated into one possible combination of the correspondent functions from Table 1 would be represented equivalently as:

Capture : Fix : Release : Convey : Control : Transform : Disperse.

There are many such possible combinations of tertiary functions that map to the abstract form. The functional modeling approach naturally provides hierarchical semantics that can associate near- and far-field concept descriptions.

1.3 Patent Datasets

Patents have been considered sources of analogies and concepts that can lead to innovative solutions [55]. In addition to the sheer volume of information contained in the patent database, all the concepts within the database must be both useful and novel. "Useful" is defined as being functional and operable, and "novel" is defined as being non-obvious and having not previously existed in the public domain [55].

Another valuable feature of the patent database for design information retrieval is the patent classification structure of the USPTO. Approximately 450 well-defined primary classification categories have been established to organize and group patents according to the field of invention. The classification system is a powerful element that benefits information retrieval by enabling data clustering for more efficient presentation and organization of search results [55]. Patents are structurally well formed with distinct partitions, and the sections that contain the embedded design

information are the abstract, claims and description. The regular structure of the documents will enable relatively simple implementation of natural language processing techniques to extract functional information. A review of patent search and information extraction literature exposed a dearth of literature on function extraction and concept generation from patents in general. Much of the literature is related to the topics of patent invalidity searches and patent informatics [56, 57], but the same information extraction principles will be applied for deriving the patent functionality.

A significant focus of the literature has been computational design aids using the patent database. TRIZ is the basis of many of these design aids. It is a theory which presents heuristic rules, or principles, to assist designers to overcome impasses in functional reasoning based on previous classification of patents in terms of contradictions [58]. Zhang et al. have used the functional basis in combination with TRIZ to create an axiomatic conceptual design model [59]. Using textual analysis of patents for use in TRIZ, Cascini and Russo presented a way to automatically identifying the contradiction underlying a given technical system [60]. To identify relevant candidates for TRIZ automatically, Souilli et al. developed a method using linguistic markers [61, 62].

Patent mining is another area of study in which meta-data is used to identify or understand large sets of patents. For example, patent citations have been used to understand the interrelatedness between patent technologies, and the benefits of understanding the pre-existing knowledge within a domain [63]. Methods have also been created to judge possible future market trends, identify prolific inventors, and more, for business applications [64, 65].

The mapping of patents has been an additional area of significant research. A method of extracting inherent structure in textual patent data has been implemented for both studying and supporting design-by-analogy [66-68]. Szykman et al. have built design repositories to share and reuse elements of designs in the development of large scale or complex engineering systems [69]. Mukherjea et al. found semantic associations between important biological terms within biomedical patents, using a semantic web with the intent of aiding in the avoidance of patent infringement [70]. Chakrabarti et al. used a topic model to analyze patent data, leading to a taxonomy or hierarchical structure [71].

While the U.S. patent database is a ripe repository to support design-by-analogy, the size and complexity make it very difficult to access in a top down way. Attempts to aid in the search and use of the patent database include theories like TRIZ and their resulting tools [58, 59, 72-82], along with many more research driven tools and methods [35, 52, 83-85]. Previous work in this field most often relies deeply on users and designers to create their own analogies, or search through large quantities of results.

In summary, there is a rich body of research on functional modeling and representation of concepts, a rich body of research on design-by-analogy, and a rich body of research on patent indexing and analysis. However, these three sets of works remain independent. We have brought these works together to apply the natural descriptive capability of functional modeling to draw analogies between functionally described design problems and functionally described patent documents. In the next section, the underlying mathematical formalism to represent functionality is reviewed, the function vector space model. Then, the modeling and matching algorithms are reviewed, followed by discussion of the finding and presentation of analogous patent documents and how they could be used within a design process. We then present a case study, and compare the method to traditional patent searching to test for efficacy.

2 METHODOLOGY AND EMBODIMENT TOOL

The development and implementation of the function vector approach to analogy search is a five-step process shown in Figure 2. It begins by constructing a controlled vocabulary of functions extracted from the patent database (i.e., mapping the general functional basis to an equivalent functional language basis for patents), making use of the hierarchical structure of the functional basis. Once a complete set of patent function terms is compiled, a basis set of the patent function terms is defined. Then, the patent documents are indexed against the functional basis to create a vector representation of the patent database. Query generation and similarity ranking tools are then developed to query and retrieve the patents with the highest degree of relevance to the functional description of a given design problem. Finally, the most relevant patent results are presented to the user. These steps are now detailed.

2.1 Knowledge Database Processing

As shown in Figure 2, the first step of the five-part process involves retrieving the design document (patent) information in the form of text, parsing that text, and then implementing tokenizing, or breaking down passages of text into their individual words or “tokens”, and word stemming, or reducing words to their base or root form. The vector space model (VSM) of information retrieval is used as the basis of the analogy search method developed in this work [86]. VSM was first developed in the early 1970’s to overcome several limitations of the Boolean model, such as lack of search result relevancy ranking, strict query syntax requirements, and query expansion limitations [1, 2]. In VSM, a document is represented as a vector of terms. The terms are words and/or phrases extracted from the documents themselves

using natural language processing techniques [87, 88]. To represent a document as a vector of terms, each term in the vocabulary becomes an independent dimension in an n -dimensional space, where n is the number of vocabulary terms. All of the documents in the database are mapped onto the vector space using indexing algorithms. In the most basic algorithm, binary values are assigned for each dimension according to whether the term occurs in the document, 1 for present and 0 for absent, but typically a weighting factor is applied to the occurring terms [86]. The two common weighting factors are the *term frequency* (tf), which is the frequency of occurrence within a specific document, and the *document frequency* (df), which is the frequency of occurrence across documents [9, 10].

The resulting term-document matrix is a matrix of size $m \times n$, where m is the number of documents in the collection, and n is the number of terms, and is typically a very sparse matrix given that relatively few terms occur within a single document. A variant of the standard VSM model called latent semantic indexing (LSI) or latent semantic analysis (LSA) can be used to reduce the dimensionality of the term-document matrix [9]. Using term co-occurrence information, singular value decomposition (SVD) methods map the document terms to a reduced *concept* space [89]. In this context, *concepts* are groups of terms that are synonyms, hypernyms, and troponyms of each other. For example, the terms *car*, *truck*, *pickup* and *automobile* are synonyms and/or hypernyms, where a hypernym is defined as a generalized term that more specific terms fall under. Troponyms apply only to verbs and are defined as verbs that more specifically describe the action. For example, *march* is a troponym of *walk*. Using SVD, the four terms can be clustered into a single dimension. Applied across the

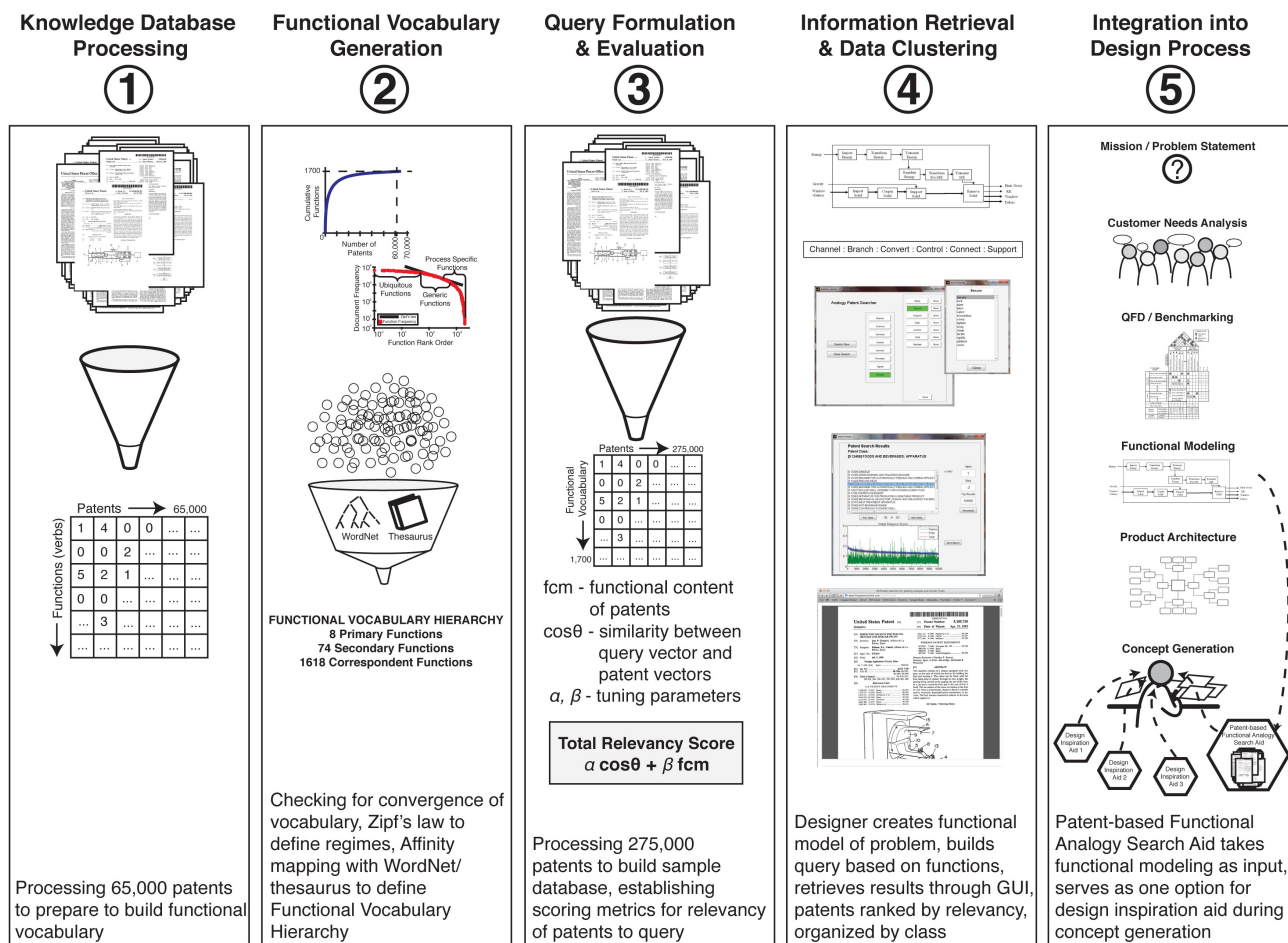


FIGURE 2: OVERVIEW OF THE FUNCTIONAL ANALOGY SEARCH DEVELOPMENT

entire term-vector, the n -dimensional space, typically in the thousands of terms, is reduced to a k -dimensional space, typically in the hundreds of concepts, and the dimensionality of k is a system parameter that must be tuned to optimize the mapping.

Some drawbacks to LSI are high computational requirements for the SVD algorithm and difficulties in adding documents to the database. Adding large numbers of new documents to a database without recomputing the SVD can lead to skewed similarity results and omitted terms. This issue is particularly significant in the patent database where documents are continually added [9]. Given the added computational overhead, issues with document additions, and marginal performance improvement, the standard VSM approach was chosen over LSI as the search engine model for this research. Issues of polysemy, where a word has multiple meanings, and synonymy, where multiple words have the same meaning, are overcome through query mapping heuristics using one-to-many term mapping; in other words, query mapping rules are devised such that a single query term is mapped to multiple document terms, allowing for the simplified query to capture a range of patents that possess the same general functionality.

One of the powerful aspects of the VSM model is that queries can be mapped to the term vector space using the same algorithms as the document mapping. This flexibility removes the syntactic constraints on the query structure and provides a simple, straight-forward metric for evaluating similarity between the query and the documents [9]. In the term-vector space, the similarity between the query vector and a document vector is equivalent to the angle between the vectors. The cosine of the angle between the vectors is a commonly used metric since it has the useful properties of varying from 0 for orthogonal vectors and 1 for identical vectors [2]. Finally, as stated above, an aspect of the VSM model that is exploited in this research is the capability to establish query mapping rules to map a single query term to multiple document terms. The ability to utilize this synonymy leads to the retrieval of a range of patents with the same general functionality. For example, if the single query term were “divide”, as shown in Table 1, synonymous terms such as “section, branch, partition, segregate, dissect, etc.” would also be included in the query.

Because purely manual indexing is very tedious and resource intensive, tools were developed to preprocess the patents using natural language processing techniques. The patent text is parsed directly from HTML to extract information, such as the title, abstract, description, claims, and patent class. Stop words lists are used to eliminate unnecessary terms, such as articles and prepositions [10]. In addition, word stemming algorithms are applied to the retrieved text to further consolidate terms. A modified Porter stemming algorithm is applied to terms to strip suffixes, e.g. -ing, -s, -es [90]. The Porter stemmer is too aggressive for the purpose of this research; for example, component-noun terms *connector* and *connection* are stemmed to the function term *connect* using Porter [90]. A modified prefix stripping algorithm was created to extract root functions. Stripped prefixes include 'sub', 're', 'un', 'de', 'under', 'mis', 'over', 'pre', 'post', 'non', 'counter', 'out', 'inter', 'micro', 'up', 'super', 'en', 'co', 'dis', 'hyper', 'ultra', 'anti'. A major component of automated indexing of the patents involves part-of-speech (POS) tagging. Here, we used TreeTagger, an open-source POS tagging program chosen based on

high accuracy of tagging in natural language documents. Tests of accuracy have shown it to be over 95% accurate [3]. TreeTagger program identifies the POS from sentence structure using probabilistic, binary decision trees [3]. Automated indexing was validated with manual verification.

2.2 Functional Vocabulary Generation

A primary goal of this research is to identify and extract a complete set of functions covering the entirety of the patent database, depicted in step 2 of Figure 2. Completeness of the function vocabulary is evaluated using two metrics: *cumulative functions versus number of patents indexed* and *function document frequency versus term chronological order*. After indexing 65,000 randomly selected patents (limited by the maximum database size), a set of approximately 1,700 functions are identified. A secondary database could be constructed to expand the capability beyond 65,000 patents if completeness has not been achieved, but this step is not necessary per the results presented next. In Figure 3, cumulative functions plotted versus patents illustrates that the metric has reached a horizontal asymptote, and furthermore convergence was reached at approximately 61,000 patents. This asymptote provides a verification that the function vocabulary does in fact converge to finite set. Therefore, any user of the methodology need not recreate this list of 1,700 patent basis functions; our one-time generation of this list suffices. On the other hand, this can be periodically rechecked easily, and is presented in detail here for scientific repeatability of the development method.

The plot in Figure 4 shows the document frequency of the function versus the order in which the function was first identified. The document frequency measures how often a term occurs across all patents. Statistically, high document-frequency terms will be found earlier due to the random sampling. The trend shown in Figure 4 is clearly confirmed with the functions' document frequencies clustering below 1% of searchable patents as a function of order found. The 1% threshold is chosen not based on a hard limit, but from the insight that terms below that level are excluded from 99% of the remaining patents. The low resolving power of these low frequency terms means little value is added to search queries by including them, since they will have no impact on similarity for the vast majority of patents. The resolution power of terms as a function of frequency is a reflection of Zipf's law [11, 12, 87], here not for all written documents, but rather only for patent documents.

The frequency of words follows a power law distribution (straight line on log-log scale) and the resolving power is analogous to a Gaussian distribution, where both very high frequency terms and very low frequency terms have low resolving power. This reasoning for the high frequency terms is the underlying justification for using the stop words lists. The upper and lower cut-offs are therefore thresholds and can be selected based on considering how many additional documents one seeks to consider versus the risk in excluding too many documents. No direct equation exists to make this determination, where others have advocated a trial and error tuning process [9]. The function vocabulary identified in the indexing process is plotted in Figure 4, using log-log axes. A Zipf distribution was fit through the data for comparative purposes, as shown in Figure 5, quantifying the resolving power of different terms.

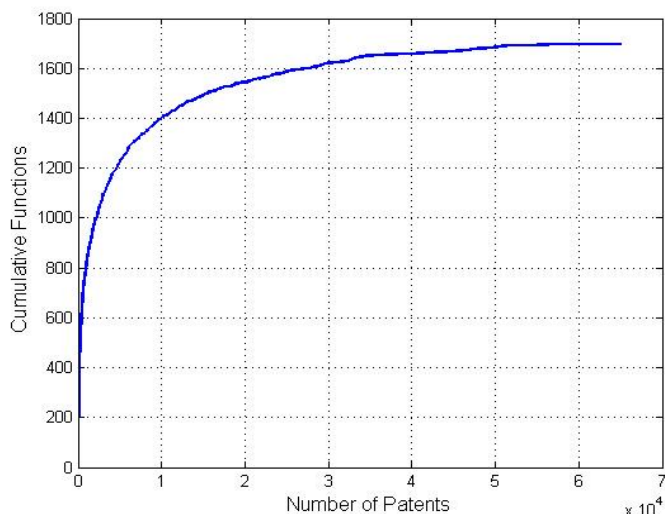


FIGURE 3: CUMULATIVE FUNCTIONS VERSUS NUMBER OF PATENTS INDEXED WITH HORIZONTAL ASYMPTOTE AT ~1,700 FUNCTIONS AND 61,000 PATENTS VERIFYING CONVERGENCE OF FUNCTION VOCABULARY

Examining Figure 5, when compared to Zipf's law, three different regimes of function frequency distribution can be identified and are labeled as: ubiquitous, generic and process-specific. Ubiquitous functions occur so frequently across all patents that they offer little value for determining similarity or relevance, per Zipf's theory. These functions can be considered to lie above the upper cut-off, chosen to be all terms that occur in more than 50% of patents. Examples of these functions are *provide*, *use*, etc. The ubiquitous functions, which account for 50 of the 1,700 terms, are to be removed from the final function vocabulary index. Generic functions have a good balance between frequency and specificity to enable better distinction between patent vectors within the cosine similarity metric. Examples of these functions are *shape*, *rotate*, etc. Process-specific functions occur in very few patents and would be below the lower cut-off region. Blindly following the resolving power hypothesis, these terms should be removed from the function index as well, but the rarity of the function may in and of itself lead to novel solutions. The retention of these few extra terms does not impact the computational overhead since the converged and complete functional vocabulary consists of just over 1,700 terms after removal of the ubiquitous functions. The patent based functional analogy search methodology can now be developed using the functional vocabulary derived in this section of work.

After the final set of functions is vetted per the process described previously, affinity diagramming and thesaurus construction techniques were used to create a hierarchical structure for the 1,700 word functional vocabulary, modeled after the functional basis [7, 8]. The affinity diagram technique is used to group like terms together into sub-groups of hypernyms and synonyms. Unusual or unfamiliar words were checked against existing thesauri to select the proper grouping. The iterative process created secondary functions with similar numbers of correspondent sub-functions. The function sub-groups were split or merged accordingly to attain consistent numbers of functions in each sub-group. The detailed procedure, all performed entirely computationally except for the use of the thesaurus and WordNet in steps 1 and 4 below, for developing the hierarchical structure of the expanded functional basis is given as follows:

1. Sort all terms into primary basis functions using thesaurus and WordNet according to synonymy and hypernym relationships [13, 14].
2. Rank verbs within each primary group by document frequency.

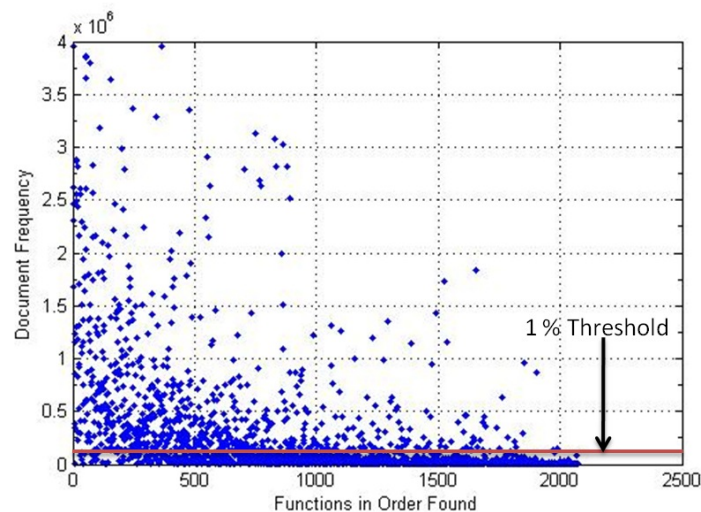


FIGURE 4: TERM DOCUMENT FREQUENCY VERSUS ORDER, SHOWING THE FREQUENCY FALLS BELOW A 1% THRESHOLD (OCCUR IN < ~4500 PATENTS)

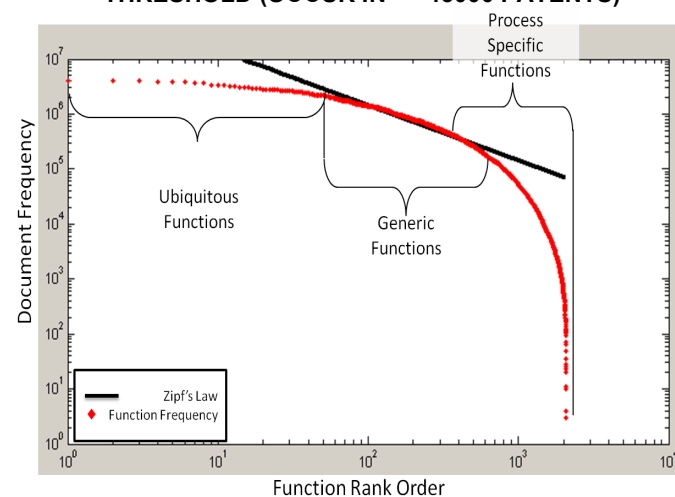


FIGURE 5: FUNCTION VOCABULARY DOCUMENT-FREQUENCY VERSUS RANK ORDER COMPARISON WITH ZIPF'S POWER LAW DISTRIBUTION

3. Review verbs and extract five highest frequency terms. These terms become initial secondary functions.
4. Group remaining correspondent functions within each secondary group using thesaurus and WordNet hierarchical relationships [13, 14].
5. Rank verbs within each secondary group by document frequency.
6. Separate groups that contain more than 50 verbs into multiple secondary function groups.
7. Iterate on grouping process to produce secondary function groups with similar number of correspondent functions.

The resulting structure of the expanded functional basis vocabulary is 1,700 unique functions organized into 74 groups of secondary functions. The secondary functions and associated correspondents are mapped into the eight (8) primary functions. Table 1 illustrates the hierarchical structure for two of the secondary functions: divide and import.

This result is readily scalable to add new patents. Utilizing the structure of the function vocabulary, a patent search sample database was constructed by indexing additional patents against the completed function vocabulary. For the purposes of this research, a representative

TABLE 1: EXAMPLES FROM THE EXPANDED FUNCTIONAL BASIS VOCABULARY FOR THE SECONDARY FUNCTIONS OF DIVIDE AND IMPORT

Primary	Secondary	Correspondents	Primary	Secondary	Correspondents
Branch	Divide	Section	Channel	Import	Permit
		Divide			Insert
		Segment			Input
		Branch			Introduce
		Sort			Inlet
		Partition			Accept
		Tab			Admit
		Miss			Fetch
		Diverge			Inflow
		Fractionate			Breathe
		Cube			Aspirate
		Segregate			Import
		Dissociate			Invite
		Graduate			Ingest
		Quantize			Invade
		Parse			Inhale
		Allot			Include
		Buck			Obtain
		Dissect			Receive
		Dismantle			Enter
		Shred			Cannulate
		Interdigitate			Induct
		Packetize			Internalize
		Compartmentalize			Imbibe
		Part			
		Separate			
		Digitalize			
		Modularize			
		Butcher			
		Sectionalize			

sample database of patents was constructed from a subset of the USPTO patent database. Three continuous selections of 100,000 patents each were chosen to be indexed. The patent groups were selected chronologically, with the first selection from patents 3,560,000 to 3,660,000, the second selection from patents 5,000,000 to 5,100,000, and the final selection from patents 7,500,000 to 7,600,000, spanning the years from 1971 to 2009. The reasoning behind this is that the creation of patents is exponentially increasing with time; so, it follows that if a random set was chosen from the entire database, many more would be from recent years than from further in the past. Choosing three sets within three ranges of patent numbers coming from three distinct bands of time was an attempt to get a more even set of patents temporally. After omitting repeated or missing patents, the sample database consists of approximately 275,000 patents mapped into document vectors, resulting in an approximately 275,000 x 1,700 patent vector matrix. The whole of the patent database was not indexed, as this was an example implementation of the methodology, in addition to the limitations of the current hardware and software prototype implementation; however, it is not unreasonable to achieve this goal in the near future.

2.3 Query Formulation and Evaluation

The next step (step 3 in Figure 2), of the research was to formulate the means to query the database of patents and functions. The binary document vector matrix contains both the functional content information for each patent as well as the term-document frequencies across all patents indexed. The term document frequency and the patent functional content are used to derive the similarity metric for ranking the search results. As discussed previously, the *document frequency* (df) is a common term weighting scheme and in particular the *inverse document frequency* (idf) is used to weight rare terms higher than common terms [9, 10]. The *inverse document frequency* is given as:

$$idf_t = \log \frac{N}{df_t} \quad (1)$$

where N is the total number of documents and df_t is the document frequency of term t . Previous research has shown more specific function verbs can yield more novel solutions [91], and the idf weighting yields a higher cosine similarity score for patents that contain process-specific functions. The idf is calculated for each term, and each element of the document vector matrix is scaled according to the calculated weight for that term. Furthermore, each document vector is normalized to generate a patent document unit vector matrix. The normalization is completed to simplify the cosine similarity calculation. The *patent functional content* (fcm) metric is a normalized measure of the total functional content with a specific patent. The equation for the fcm metric is given as:

$$fcm_k = \frac{\text{total number of terms in patent}_k}{\text{total number of terms in database}} \quad (2)$$

The fcm metric increases the weighting of patents with high functional content. The reasoning for including this metric is a hypothesis that functionally rich patents, or those which contain a large number of functional terms and thus explicitly address more functionalities, contain more information that can be mapped as analogies. The *total relevancy score* is then defined as a linear combination of the two components: the idf -weighted cosine similarity and the *patent functional content* metrics, summarized in Table 2.

The linear combination within the total relevancy score is weighted with two coefficients, alpha, α , and beta, β . These coefficients are tuning parameters used to bias the relevancy ranking

TABLE 2: METRICS FOR CALCULATING SIMILARITY BETWEEN THE DOCUMENT AND QUERY VECTORS

- Query-Patent Cosine Similarity	$\cos \theta = \frac{\text{Query} \cdot \text{Patent}}{\ \text{Query} \ * \ \text{Patent} \ }$
- Patent Functional Content	$FCM = \frac{\sum \text{Patent}_{term(t)}}{\text{NumTerms}}$
- Total Relevancy Criteria	$Score = \alpha \cdot \cos \theta + \beta \cdot FCM$

towards a higher weighting on either the cosine similarity or the functional content metric. The tuning parameter weights were explored empirically through a parametric evaluation process by running multiple patent searches and finding values that produced patents sufficiently near- and far-field.

To do this, a *Query Generator Tool* was created to automate the process of constructing the patent query vector. The GUI builds the query using the expanded functional basis vocabulary hierarchical structure. First, as shown in Figure 6, the user selects the primary high level function corresponding to the high level functionality derived from the functional model of the design problem. Next, the user selects one of many secondary functions, which are more detailed versions of the primary function, corresponding to the specific functionality that will be retrieved. Once the secondary function is selected, the interface populates the query vector with all correspondent terms associated within the secondary function. Additional secondary functions can then be selected to further populate the query vector for a particular primary function. The new query vector is then saved once all secondary functions are chosen. The process can then be repeated for additional primary functions. An example of functional modeling of a design problem and the subsequent primary and secondary functional term selections are detailed in Section 3.

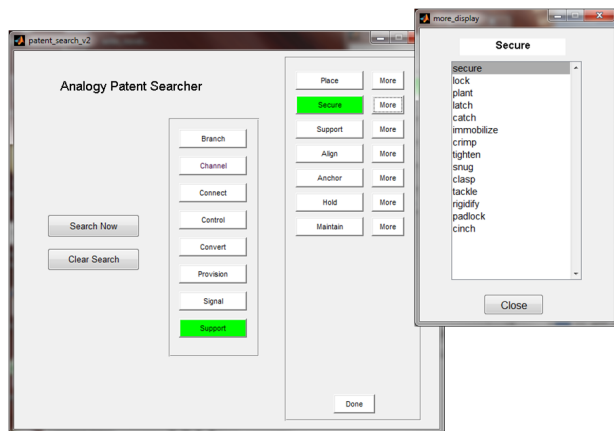


FIGURE 6: QUERY GENERATOR USER INTERFACE

2.4 Information Retrieval and Data Clustering

Once the query construction is complete, the information retrieval and clustering task is next needed, shown as step 4 in Figure 2. This is implemented in a *Search Result Viewer*, shown in Figure 7. The viewer performs multiple functions including calculating the cosine similarity, fcm , and total relevancy score, extracting the top results and clustering the results by patent class. The cosine similarity is calculated for all documents simultaneously by first normalizing the query vector to form the query unit vector, and then calculating the dot product of the unit query vector with the document vector unit matrix using the equation:

$$\overrightarrow{COS}_{\text{similarity}} = \mathbf{q}^T \cdot \mathbf{d} \quad (3)$$

where $\overrightarrow{COS}_{\text{similarity}}$ is a vector containing all cosine similarity scores for the dot product of the query vector, q , and the document vector matrix, d . The total relevancy vector is calculated by the linear sum of the COS_{vector} and the functional content metric vector, weighted by the user-defined α and β coefficients respectively. The top n results as specified by the user are retrieved, sorted by total relevancy score and clustered by primary patent classification. As shown in Figure 7, the similarity scores for the individual patents are clearly indicated in the first column of the results list. The average relevancy score for the patent class is given before the title to help the user quickly identify patent classes with high potential for identifying functionally relevant patents.

Selecting one of the search results automatically opens a web browser window with a PDF version of the selected patent, by making calls to online patent databases and using their patent viewer. The PDF version is displayed due to the fact the patent illustrations are included, as opposed to the text-only version of the patent.

To determine the optimal weighting for the total relevancy score coefficients, several searches were conducted over various function combinations. The *Search Result Viewer* interface enables the coefficients to be varied in real-time for the same search query, allowing for multiple iterations for the same function query. Following a trial-and-error process where β is varied from 1 to -1 keeping $\alpha = 1$, the search results provided more functionally relevant results for negative values of the fcm coefficient β . This result contradicts the thought that functionally rich patents are more readily mappable to functional analogies. We found the fcm metric not as useful as it intuitively appears. Patents with high fcm were thought to contain a high percentage of function terms. In practice, however, instead, positive values of β skew the results towards long patents since,

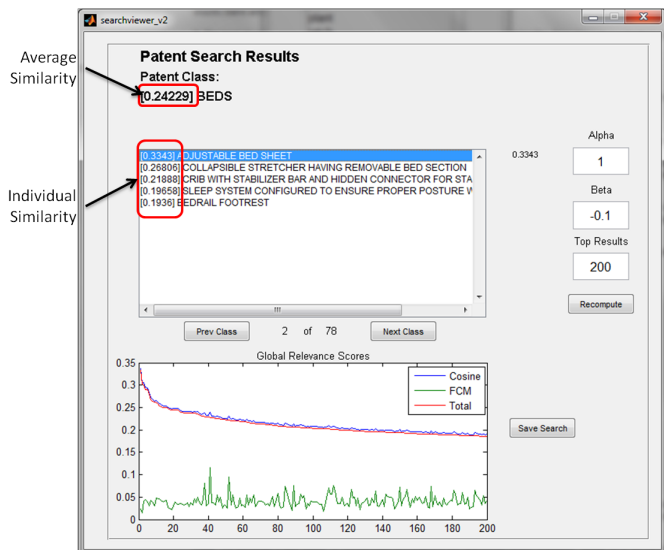


FIGURE 7: SEARCH RESULT VIEWER SHOWING AVERAGE TOTAL RELEVANCY SCORE FOR PATENT CLASS AND INDIVIDUAL TOTAL RELEVANCY SCORE FOR IDENTIFIED PATENTS

statistically, patents that contain more text will contain more function verbs. Elucidating useful analogies from these broad patents is cognitively more difficult than functionally focused patents. Therefore, empirically, the default values for α and β are set to 1 and -0.2, respectively, which focused the total relevancy score toward functionally focused patents.

2.5 Integration into Design Process

The last step of the method (step 5 in Figure 2) is to make use of the resulting patents presented. The steps described in Sections 2.1-2.4 are combined into a structured methodology for identifying analogous patents. With the concept generation process, the analogy search methodology is used as a supplemental technique to more traditional concept generation methods, such as brainstorming, brainsketching, and the CSketch/6-3-5 method [8, 92-94]. Device functionality developed early in a functional modeling phase can be used directly to create functional semantic representations of the design problem by simply stripping the verbs from the functional description. These function verbs can then be mapped to the primary and secondary functions through the expanded functional basis vocabulary. The *Query Generation Tool* can then be utilized to create the query function vector for the device. The *Search Result Viewer* algorithms identify the functionally similar patents in which analogies to the design problem likely exist. Then the user can review these sorted patents and consider them for analogical solutions back to the original problem domain. To consider the efficacy of this approach and others, the function analogy search methodology above is applied to a case study problem, and compared against the more traditional approach of simply using keyword patent searches.

3 EXPERIMENTAL VALIDATION OF CONCEPTUAL DESIGN PHASE EFFICACY

The case study utilized to evaluate the methodology presented in this paper is the design of an automated window washing device. The problem is to design a self-contained window cleaning device. Once initialized, the device will begin an automated routine for removing dirt, film, and debris from the window surface without user interaction. The general problem statement allows for multiple process choices

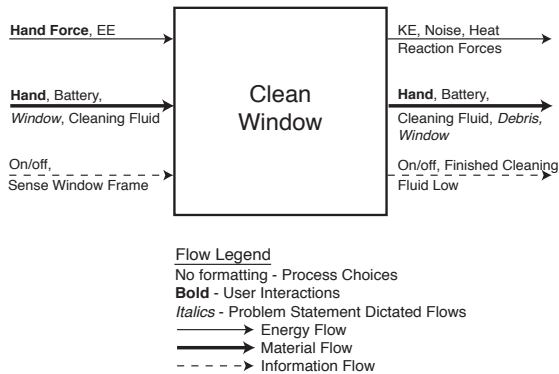


FIGURE 8: BLACK BOX FUNCTIONAL MODEL (TOP) AND SIMPLIFIED FUNCTIONAL MODEL (BOTTOM) OF CORE FUNCTIONALITY FOR AN AUTOMATED WINDOW WASHER

such as the power source and cleaning method. The black box functional model and the more simplified functional model showing core functionality for a battery-powered device that utilizes a liquid media for cleaning are shown in Figure 8. Other alternative process choices for a power source are solar-power and fuel cells, among others. Alternative cleaning method process choices omit the cleaning fluid and rely on mechanical or other energy-domain removal of debris.

The functional semantic representation of the simplified model becomes:

Import: Transform: Transmit: Regulate: Couple: Support: Remove

Further generalizing the model into the primary functions results in the functional semantic representation given as:

Channel : Branch : Convert : Control : Connect : Support

A separate analogy search is performed by the first author for

each primary function using the secondary functions most relevant to the original design problem. The multiple search approach is used to maximize the relevancy score resolution for each query. The secondary functions utilized for each search query are:

- Channel → Import, Transmit and Translate
- Branch → Remove, Clean and Disperse
- Convert → Transform and Treat
- Control → Control and Adjust
- Connect → Connect, Mount, and Couple
- Support → Secure and Align

All searches are performed using the default values for the total relevancy score metric of $\alpha = 1$ and $\beta = -0.2$. The top 500 results are retrieved for each search.

The 4th patent identified for the window cleaning device (Patent 5,086,533) is a very near-field analogy to the proposed design problem. The device shown in Figure 9a utilizes a squeegee mechanism with a fluid application system to automatically clean windows. A second cleaning device, shown in Figure 9b, is used for automatically cleaning floors.

The second patent identified is a floor cleaning robot (Patent 6,883,201) solution, better known as the iRobot Roomba™, performs the same desired functionality as the automated window washer, but the application is in a different domain (floors versus windows). Therefore, this solution is a far-field analogy that is readily adaptable to the window cleaning domain. The missing functionality of coupling the device to a window can be derived from other far-field analogies such as the 8th patent identified, a wafer polishing patent (Patent 7,559,825), which utilizes vacuum to couple the device to the wafer surface. A purely mechanical means of traversing vertical surfaces is described in Patent 5,033,586 for a transportable construction elevator, shown in Figure 9c, using a pulley mechanism.

Finally, entirely novel methods of cleaning surfaces are identified using the patent based functional analogy search methodology. The 6th patent identified, Patent 5,025,632, describes an innovative process for cleaning surfaces utilizing a combination of cryogenically cooled fluids and mechanical abrasion. Although the cryogenic solution may not be feasible in applications of cleaning glass surfaces, the purpose of the tool is to stimulate novel problem solving by identifying both near- and far-field analogies. The case study applied the search methodology to the design problem of the automated window washer.

TABLE 3: GOOGLE PATENT SEARCH RESULTS FOR “WINDOW CLEANER”

#	Patent	Title	What is it?
1	WO2013005937A3	Double-sided glass window cleaner	Two pads on a scissor linkage squirting water to hydroplane
2	CN202526075U	Convenient window cleaner	Gripper of a sponge
3	CN 203029128 U	PVC window cleaner with a suction	Plunger squirter
4	CN 203000787 U	Cleaner for plastic-steel window	Shovel shape below wiper
5	WO 2000068354 A1	Window cleaner tablet	Water-soluble effervescent cleaner pill to drop in water
6	DE 102012204028 A1	Window cleaning system	Automotive window wiper
7	US 8311440 B2	Laser scanner window cleaner	Print head wiper of the laser head
8	WO 1984003459 A1	Window cleaner	Porous dry sponge with impregnated particles
9	WO 2013069986 A1	Window cleaning apparatus capable of manipulation via magnetic attraction and control method thereof	Electromagnets on both sides of window to mount tooling
10	EP 0673992 A2	Concentrated liquid glass and window cleaning composition and method of use	Soap
11	US 3342740 A	Window cleaner	Soap
12	CN 101703382 B	Hydraulic lifting mechanism of window cleaning equipment	Elevator machine as a post and cable

Six individual searches are performed and the compiled results include both near- and far-field analogies. Among the far-field analogies are novel solutions for coupling the device to vertical surfaces using vacuum or transportable pulley systems and for removing debris using cryogenic fluids. The case study performed utilizing the analogy-based search engine shows that both near- and far-field analogies can be quickly interpreted and derived from the patents obtained.

A separate publication presents results of an experiment designed to elucidate the effects of presenting functionally analogous patents, identified using the method developed in this paper, during concept generation on the quantity and novelty of design solutions. For details of the study, the reader is referred to [95, 96].

A point of comparison for our original hypothesis is to compare these results with a simple patent search of key terms. Using the Google Patent search engine, the keywords “window cleaner” were searched. The top 12 patents are shown in Table 3, from a list of the top 500. One can see that within the 12 patents shown, 25% are unrelated soap compounds, and 17% are unrelated lift mechanisms. Only two patents are wipers, one near-field and one far-field. The two remaining patents are sponge concepts. Overall, 8 of the top 12 returned patents are unrelated. We find the functional vector approach to analogical patent search more effective.

4 CONCLUSIONS

The patent-based functional analogy search methodology provides an organized method for identifying functionally similar patents independent of the patent solution domain. The domain-independent search capability is achieved through the systematic derivation of a complete functional vocabulary extracted from the target knowledge base of the USPTO patent database. Several natural language processing algorithms are developed and implemented to identify a finite set of function verbs, and the functions are organized into an expanded functional basis vocabulary with a hierarchical structure. The 1,700 function terms are utilized to generate a searchable document vector matrix consisting of approximately 275,000 patents. Search interfaces were created to enable effortless access to the vast design information contained in the limited sample of the patent database. Additional insight gained in the model development is the knowledge that patents that are longer are more difficult to map analogically due to the longer list of functional verbs. Further research is needed to verify the optimal values for the total relevancy score metric to perhaps include patent document length. Another proposed extension to the search methodology implementation is applying the method ultimately to analogical search across large-scale and less structured data, such as the world wide web.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Matthew Campbell, Dr. Richard Crawford, and Dr. Joseph Beaman for their discussions and insights on this work. This work was supported by the National Science Foundation, grant numbers CMMI-0855326, CMMI-0855510, and CMMI-0855293, and the SUTD-MIT International Design Centre (IDC) (<http://www.sutd.edu.sg/idc.aspx>). Any opinions, findings, or recommendations are those of the authors and do not necessarily reflect the views of the sponsors.

REFERENCES

[1] G. Salton, *The SMART Retrieval System—Experiments in Automatic Document Retrieval*. Englewood Cliffs, NJ: Prentice Hall Inc., 1971.
 [2] G. Salton, A. Wong, and C. S. Yang, "A Vector Space Model for Information Retrieval," *Communications of the ACM*, vol. 18, pp. 613-620, 1975.
 [3] H. Schmid, "Probabilistic Part-of-Speech Tagging Using Decision Tree," presented at the International Conference on New Methods in Language Processing, Manchester, UK, 1994.

[4] M. Kurfman, J. Rajan, R. Stone, and K. Wood, "Functional Modeling Experimental Studies" presented at the ASME Design Engineering Technical Conferences, Pittsburgh, PA., 2001.
 [5] M. Kurfman, J. Rajan, R. Stone, and K. Wood, "Experimental Studies Assessing the Repeatability of a Functional Modeling Derivation Method," *Journal of Mechanical Design*, vol. 125, pp. 682-693, 2003.
 [6] R. Stone and K. L. Wood, "Development of a Functional Basis for Design," *Journal of Mechanical Design*, vol. 122, 2000.
 [7] J. Hirtz, Stone, R. B., McAdams, D. A., Szykman, S., Wood, K. L., "A Functional Basis for Engineering Design: Reconciling and Evolving Previous Efforts," *Research in Engineering Design* vol. 13, pp. 65-82, 2002.
 [8] K. Otto and K. Wood, *Product Design Techniques in Reverse Engineering and New Product Development*, Upper Saddle River, New Jersey: Prentice Hall, 2001.
 [9] C. D. Manning, P. Raghavan, and H. Schutz, *An Introduction to Information Retrieval*. Cambridge, England: Cambridge University Press, 2009.
 [10] G. Salton and R. K. Waldstein, "Term Relevance Weights in On-line Information Retrieval," *Information Processing & Management*, vol. 14, pp. 29-35, 1978.
 [11] R. E. Wyllys, "Empirical and Theoretical Bases of Zipf's Law," *Library Trends*, vol. 30, pp. 53-64, 1981.
 [12] G. K. Zipf, *Human Behavior and the Principle of Least Effort*: Addison-Wesley, 1949.
 [13] C. Fellbaum, *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press, 1998.
 [14] G. A. Miller, "WordNet: A Lexical Database for English," *Communications of the ACM*, vol. 38, pp. 39-41, 1995.
 [15] G. Pahl and W. Beitz, *Engineering Design: A Systematic Approach*, 2nd Edition ed. London, UK: Springer-Verlag, 1996.
 [16] J. Chan, K. Fu, C. Schunn, J. Cagan, K. Wood, and K. Kotovsky, "On the benefits and pitfalls of analogies for innovative design: Ideation performance based on analogical distance, commonness, and modality of examples," *Journal of Mechanical Design*, vol. 133, 2011.
 [17] J. Linsey, Murphy, J., Markman, A., Wood, K. L., Kortoglu, T., "Representing Analogies: Increasing the Probability of Innovation," presented at the ASME International Design Theory and Method Conference, Philadelphia, PA, 2006.
 [18] A. Chakrabarti and T. P. Blich, "A Scheme for Functional Reasoning in Conceptual Design," *Design Studies*, vol. 22, pp. 493-517, 2001.
 [19] R. Stone and K. Wood, "A Heuristic Method for Identifying Modules for Product Architectures," *Design Studies*, vol. 21, 2000.
 [20] M. Salonen, K. Holta-Otto, and K. Otto, "Effecting product reliability and life cycle costs with early design phase product architecture decisions," *International Journal of Product Development*, vol. 5, pp. 109-124, 2008.
 [21] M. R. Bohm, J. P. Vucovich, and R. B. Stone, "Capturing Creativity: Using a Design Repository to Drive Concept Innovation," presented at the Proceedings of DETC2005, DETC05/CIE-85105, Long Beach, California, 2005.
 [22] S. Szykman, R. D. Sriram, C. Bochenek, and J. Racz, "The NIST Design Repository Project," in *Advances in Soft Computing – Engineering Design and Manufacturing*, ed London: Springer-Verlag, 1999.
 [23] S. Szykman, R. D. Sriram, C. Bochenek, and J. Senfaute, "Design Repositories: Next-Generation Engineering Design Databases," *IEEE Intelligent Systems and Their Applications*, vol. 15, pp. 48-55, 2000.
 [24] C. Bryant, R. Stone, D. McAdams, T. Kurtoglu, and M. Campbell, "Concept Generation from the Functional Basis of Design," presented at the International Conference on Engineering Design, Melbourne, Australia, 2005.
 [25] C. Bryant, R. Stone, D. McAdams, T. Kurtoglu, and M. Campbell, "A Computational technique for Concept Generation," presented at the ASME Design Engineering Technical Conference, Long Beach, CA, 2005.
 [26] S. Potter, S. J. Culley, M. J. Darlington, and P. K. Chawdhry, "Automatic Conceptual Design Using Experience-derived Heuristics," *Research in Engineering Design*, vol. 14, pp. 131-144, 2003.
 [27] M. Chiu, "Design Moves in Situated Design with Case-based Reasoning," *Design Studies*, vol. 24, 2003.
 [28] D. Gentner, "Structure Mapping – A Theoretical Framework," *Cognitive Science*, vol. 7, 1983.
 [29] B. F. Falkenhainer, K. D. Forbus, and D. Gentner, "The Structure Mapping engine: Algorithm and Examples," *Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM*, vol. 41, pp. 1-63, 1989.
 [30] D. Gentner and A. B. Markman, "Structure mapping in analogy and similarity," *American Psychologist*, vol. 52, pp. 45-56, 1997.
 [31] V. V. Krysanov, H. Tamaki, and S. Kitamura, "Understanding Design Fundamentals: How Synthesis and Analysis Drive Creativity, Resulting in Emergence," *Artificial Intelligence in Engineering*, vol. 15, 2001.
 [32] G. Goldschmidt and M. Weil, "Contents and Structure in Design Reasoning," *Design Issues*, vol. 14, pp. 85-100, 1998.
 [33] J. Linsey, I. Tseng, K. Fu, J. Cagan, K. Wood, and C. Schunn, "A study of design fixation, its mitigation and perception in engineering design faculty," *ASME Journal of Mechanical Design*, vol. 132, pp. 1041003-1-12, 2010.
 [34] E. Hacco, Shu, L. H., "Biomimetic Concept Generation Applied to Design for Remanufacture," presented at the ASME Design Engineering Technology Conference and Computers and Information in Engineering Conference, 2002.

- [35] A. Goel, S. Bhatta, and E. Stroulia, "Kritik: An early case-based design system," in *Issues and Applications of Case-Based Reasoning in Design*, M. Maher and P. Pu, Eds., ed Mahwah, NJ: Erlbaum, 1997, pp. 87-132.
- [36] A. Goel and S. Bhatta, "Use of Design Patterns in Analogy-Based Design," *Advanced Engineering Informatics*, vol. 18, pp. 85-94, 2004.
- [37] D. Navinchandra, K. P. Sycara, and S. Narasimhan, "Behavioral Synthesis in CADET, a Case-Based Design Tool," presented at the Seventh IEEE Conference on Artificial Intelligence Applications, 1991.
- [38] L. Qian and J. S. Gero, "A design support system using analogy," in *Artificial Intelligence in Design*, ed: Springer Netherlands, 1992, pp. 795-813.
- [39] C. T. Charlton and K. M. Wallace, "A web broker for component retrieval in mechanical engineering," *Design Studies*, vol. 21, pp. 167-186, 2000.
- [40] Y.-C. Liu, A. Chakrabarti, and T. P. Bligh, "A computational framework for concept generation and exploration in mechanical design," *Artificial Intelligence in Design*, pp. 499-519, 2000.
- [41] Y.-C. Liu, T. P. Bligh, and A. Chakrabarti, "Towards an 'ideal' approach for concept generation," *Design Studies*, vol. 24, pp. 341-355, 2003.
- [42] A. Chakrabarti, P. Sarkar, B. Leelavathamma, and B. S. Nataraju, "A Functional Representation for Aiding in Biomimetic and Artificial Inspiration of New Ideas," *Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM*, vol. 19, pp. 113-132, 2005.
- [43] A. Chakrabarti, "Design Creativity Research," in *Product Research*, ed: Springer Netherlands, 2009, pp. 17-39.
- [44] M. C. Yang, W. H. Wood, and M. R. Cutkosky, "Design Information Retrieval: A Thesauri-base Approach for Reuse of Informal Design Information," *Engineering with Computers*, pp. 177-192, 2005.
- [45] W. H. Wood, M. C. Yang, M. R. Cutkosky, and A. M. Agogino, "Design Information Retrieval: Improving Access to the Informal Side of Design," presented at the ASME Design Engineering Technical Conference, Atlanta, Georgia, 1998.
- [46] M. C. Yang and M. R. Cutkosky, "Automated Indexing of Design Concepts for Information Management," presented at the International Conference on Engineering Design, Tampere, 1997.
- [47] S. Ahmed, "An Approach to Assist Designers with their Queries and Designs," presented at the ASME Design Engineering Technical Conference, Pittsburgh, PA, 2005.
- [48] J. Linsey, A. B. Markman, and K. L. Wood, "WordTrees: A method for design-by-analogy," presented at the ASEE Annual Conference, 2008.
- [49] J. Linsey, A. B. Markman, and K. L. Wood, "Design by Analogy: A Study of the WordTree Method for Problem Re-Representation," *Journal of Mechanical Design*, 2012.
- [50] J. S. Linsey, K. L. Wood, and A. B. Markman, "Increasing Innovation: Presentation and Evaluation of the WordTree Design-by-Analogy Method," presented at the Proceedings of the 2008 IDETC/CIE Conference, 2008.
- [51] N. M. Segers, B. De Vries, and H. H. Achten, "Do word graphs stimulate design?," *Design Studies*, vol. 26, pp. 625-647, 2005.
- [52] P. Verhaegen, J. D'hondt, D. Vandenneve, S. Dewulf, and J. R. Duflou, "Identifying Candidates for Design-by-Analogy," *Computers in Industry*, vol. 62, pp. 446-459, 2011.
- [53] D. Gentner, "Some Interesting Differences between Verbs and Nouns," *Cognition and Brain Theory*, vol. 4, pp. 161-178, 1981.
- [54] I. Chiu, Shu, L. H., "Using Language as Related Stimuli for Concept Generation," *Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM*, vol. 21, pp. 103-121, 2007.
- [55] I. Kang, S. Na, J. Kim, and J. Lee, "Cluster-based Patent Retrieval," *Information Processing and Management*, vol. 43, 2007.
- [56] A. J. Tripppe, "Patinformatics: Tasks to Tools," *World Patent Information*, vol. 25, 2003.
- [57] Y. Tseng, C. Lin, and Y. Lin, "Text Mining Techniques for Patent Analysis," *Information Processing and Management*, vol. 43, 2007.
- [58] G. S. Altshuller and R. B. Shapiro, "О Психологии изобретательского творчества (On the psychology of inventive creation)(in Russian)," *Вопросы Психологии (The Psychological Issues)*, vol. 6, pp. 37-39, 1956.
- [59] R. Zhang, J. Cha, and Y. Lu, "A Conceptual Design Model Using Axiomatic Design, Functional Basis and TRIZ," presented at the Proceedings of the 2007 IEEE IEEM, 2007.
- [60] G. Cascini and D. Russo, "Computer-aided analysis of patents and search for TRIZ contradictions," *Int. J. of Product Devel.*, vol. 4, pp. 52-67, 2007.
- [61] A. Souili, D. Cavallucci, F. Rousselot, and C. Zanni, "Starting from patents to find inputs to the Problem Graph model of IDM-TRIZ," presented at the TRIZ Future 2011, Dublin, Ireland, 2011.
- [62] A. Souili and D. Cavallucci, "Toward an automatic extraction of IDM concepts from patents," presented at the CIRP Design, 2012.
- [63] A. K. Chakrabarti, I. Dror, and E. Nopphadol, "Interorganizational Transfer of Knowledge: An analysis of Patent Citations of a Defense Firm," *IEEE Transactions on Engineering Management*, vol. 40, pp. 91-94, 1993.
- [64] C. Magee, "Towards quantification of the role of materials innovation in overall technological development," *Complexity*, vol. 18, pp. 10-25, 2012.
- [65] C. L. Benson and M. C. L., "A hybrid keyword and patent class methodology for selecting relevant sets of patents for a technological field," *Scientometrics*, vol. 96, pp. 69-82.
- [66] K. Fu, "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, USA, 2012.
- [67] K. Fu, J. Cagan, K. Kotovsky, and K. Wood, "Discovering Structure in Design Databases Through Function and Surface Based Mapping," *Journal of mechanical Design, In Press*, 2013.
- [68] K. Fu, J. Chan, J. Cagan, K. Kotovsky, C. Schunn, and K. Wood, "The Meaning of 'Near' and 'Far': The Impact of Structuring Design Databases and the Effect of Distance of Analogy on Design Output," *ASME Journal of Mechanical Design*, vol. 135, p. 021007, 2013.
- [69] S. Szykman, R. D. Sriram, C. Bochenek, J. W. Racz, and J. Senfaute, "Design Repositories: Engineering Design's New Knowledge Base," *IEEE Intelligent Systems*, pp. 48-55, 2000.
- [70] S. Mukherjee, B. Bhuvan, and P. Kankar, "Information Retrieval and Knowledge Discovery Utilizing a BioMedical Patent Semantic Web," *IEEE Transactions on Knowledge and Data Engineering* vol. 17, pp. 1099-1110, 2005.
- [71] S. Chakrabarti, B. Dom, R. Agrawal, and P. Raghavan, "Scalable Feature Selection, Classification and Signature Generation for Organizing Large Text Databases into Hierarchical Topic Taxonomies," *The VLDB Journal* vol. 7, pp. 163-178, 1998.
- [72] R. Duran-Novoa, N. Leon-Rovira, H. Aguayo-Tellez, and D. Said, "Inventive Problem Solving Based on Dialectical Negation, Using Evolutionary Algorithms and TRIZ Heuristics," *Computers in Industry*, vol. 62, pp. 437-445, 2011.
- [73] N. V. Hernandez, L. C. Schmidt, and G. E. Okudan, "Systematic Ideation Effectiveness Study of TRIZ," presented at the ASME IDETC/CIE, Chicago, IL, USA, 2012.
- [74] N. V. Hernandez, L. C. Schmidt, and G. E. Okudan, "Experimental Assessment of TRIZ Effectiveness in Idea Generation," presented at the ASEE Annual Conference, San Antonio, TX, USA, 2012.
- [75] V. Krasnoslobodtsev and R. Langevin, "TRIZ Application in Development of Climbing Robots," presented at the First TRIZ Symposium, Japan, 2005.
- [76] Y. Liang, R. Tan, and J. Ma, "Patent Analysis with Text Mining for TRIZ," presented at the IEEE ICMIT, Bangkok, Thailand, 2008.
- [77] T. Nakagawa, "Creative Problem-Solving Methodologies TRIZ/USIT: Overview of my 14 Years in Research, Education, and Promotion," *The Bulletin of the Cultural and Natural Sciences in Osaka Gakuin University*, vol. 64, March 2012.
- [78] A. A. Nix, B. Sherret, and R. B. Stone, "A Function Based Approach to TRIZ," presented at the ASME IDETC/CIE, Washington, D.C., USA, 2011.
- [79] R. Houssin and A. Coulibaly, "An Approach to Solve Contradiction Problems for Safety Integration in Innovative Design Process," *Computers in Industry*, vol. 62, pp. 398-406, 2011.
- [80] D. Mann, S. Dewulf, B. Zlotin, and A. Zusman, *Matrix 2003, Updating the TRIZ Contradiction Matrix*. Belgium: CREAM Press, 2003.
- [81] CREAM. (7 September 2012). *CREAX: Creativity for Innovation*. Available: <http://www.cream.com>
- [82] I. M. Goldfire. (19 February 2012). *Invention Machine Goldfire: Unleashing the Power of Research*. Available: <http://inventionmachine.com/products-and-services/innovation-software/goldfire-research/>
- [83] S. Bhatta and A. Goel, "From design experiences to generic mechanisms: model-based learning in analogical design," *AIEDAM*, vol. 10, pp. 131-136, 1996.
- [84] I. Chiu and L. H. Shu, "Bridging cross-domain terminology for biomimetic design," in *ASME IDETC/CIE*, Long Beach, CA, USA, 2005.
- [85] J. F. V. Vincent, O. A. Bogatyreva, N. R. Bogatyreva, A. Bowyer, and A. K. Pahl, "Biomimetics: its practice and theory," *Journal of the Royal Society Interface* vol. 3, pp. 471-482, 2006.
- [86] G. Salton and M. J. McGill, *Introduction to Modern Information Retrieval*. New York, NY, USA: McGraw-Hill, Inc., 1986.
- [87] C. J. van Rijbergen, *Information Retrieval*. Oxford, UK: Butterworth-Heinemann, 1979.
- [88] T. C. Rindfleisch, "Natural Language Processing," *Annual Review of Applied Linguistics*, vol. 16, 1996.
- [89] A. Moldovan, R. I. Bot, and G. Wanka, "Latent Semantic Indexing for Patent Documents," *International Journal of Applied Mathematics and Computer Science*, vol. 15, pp. 551-560, 2005.
- [90] M. F. Porter, "An Algorithm for Suffix Stripping," *Program*, vol. 14, pp. 130-137, 1980.
- [91] J. Linsey, Claus, E. F., Kurtoglu, T., Murphy, J. T., Wood, K. L., Markman, A. B., "An Experimental Study of Group Idea Generation Techniques: Understanding the Roles of Idea Representation and Viewing Methods," *Journal of Mechanical Design*, vol. 133, p. 031008, 2011.
- [92] A. Osborn, *Applied Imagination*. New York, NY: Scribner, 1957.
- [93] A. B. Markman, Wood, K. L., *Tools for Innovation: The Science Behind Practical Methods that Drive New Ideas*. New York, NY: Oxford University Press, 2009.
- [94] A. B. Vangundy, *Techniques of Structured Problem Solving*, 2nd Edition ed. NY: Van Nostrand Reinhold Company, 1988.
- [95] J. T. Murphy, "Patent-based analogy search tool for innovative concept generation," Ph.D. Dissertation, Department of Mechanical Engineering, The University of Texas, Austin, TX, 2011.
- [96] K. Fu, J. Murphy, M. Yang, K. Otto, D. Jensen, and K. L. Wood, "Investigating the Effect of Functionality Level of Analogical Stimulation on Design Outcomes," presented at the Design Engineering Workshop, Kitakyushu, Fukuoka, Japan, 2013.