

**Jeremy Murphy**  
Schlumberger Limited,  
Sugarland, TX 77478

**Katherine Fu<sup>1</sup>**  
Massachusetts Institute of Technology,  
Singapore University of Technology and Design,  
Cambridge, MA 02139

**Kevin Otto**  
Engineering Product,  
Development Pillar,  
Singapore University of Technology and Design,  
138682, Singapore

**Maria Yang**  
Massachusetts Institute of Technology,  
Cambridge, MA 02139

**Dan Jensen**  
United States Air Force Academy,  
Colorado Springs, CO 80840

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

# Function Based Design-by-Analogy: A Functional Vector Approach to Analogical Search

*Design-by-analogy is a powerful approach to augment traditional concept generation methods by expanding the set of generated ideas using similarity relationships from solutions to analogous problems. While the concept of design-by-analogy has been known for some time, few actual 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 functional analogies from data sources has been developed to provide this capability, here based on a functional basis rather than form or conflict descriptions. Building on past research, we utilize a functional vector space model (VSM) to quantify analogous similarity of an idea's functionality. We quantitatively evaluate the functional similarity between represented design problems and, in this case, patent descriptions of products. We also develop document parsing algorithms to reduce text descriptions of the data sources down to the key functions, for use in the functional similarity analysis and functional vector space modeling. To do this, we apply Zipf's law on word count order reduction to reduce the words within the documents down to the applicable functionally critical terms, thus providing a mapping process for function based search. The reduction of a document into functional analogous words enables the matching to novel ideas that are functionally similar, which can be customized various ways. This approach thereby provides relevant sources of design-by-analogy inspiration. As a verification of the approach, two original design problem case studies illustrate the distance range of analogical solutions that can be extracted. This range extends from very near-field, literal solutions to far-field cross-domain analogies.*

[DOI: 10.1115/1.4028093]

## 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 nonobvious 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 nonobvious 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].

Pahl and Beitz created the hierarchical structure of the *functional basis* and the five categorical functions of which all functions are more specific instances [15]. Building off of this foundational work, 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 [8]. The black box model is then decomposed into subfunctions 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].

Extensive efforts have produced a standard language for representing functions and flows associated with each subfunction called the *functional basis* [6,7]. In particular, Cheong et al. and Shu et al. have made significant contributions to the field of biomimetic design and design-by-analogy through finding biologically meaningful keywords that correspond to the functional basis [19], as well as using natural language techniques to extract functionally relevant information and inspiration from biology texts [20]. 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 into 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.

<sup>1</sup>Corresponding author.

Contributed by the Design Theory and Methodology Committee of ASME for publication in the JOURNAL OF MECHANICAL DESIGN. Manuscript received June 20, 2013; final manuscript received July 16, 2014; published online August 14, 2014. Assoc. Editor: Jonathan Cagan.

Functional modeling is also used to identify modules and interface boundaries, such as that devised by Stone and Wood to transform product function into alternative product layouts by identifying modules for modular product architectures [21]. 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 Ref. [22].

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 [23–27]. In addition, computational tools have been developed to exploit the knowledge contained in the design repositories for the purpose of concept variant generation [28–31]. 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]. 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 [32–34]. 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 [16,35]; this can be conceptualized as drawing inspiration from a domain of expertise or application that is unrelated to the target domain, such as using biological phenomena to inspire the design of mechanical devices. In general, it is rare to find a designer with expertise in more than one technical field of application, making designing in this manner difficult to perform. The literature on design-by-analogy, distance has been defined thus far in a relatively qualitative and subjective manner—“near-field” being when the target and source of the analogy originate from the same domain of application, and “far-field” being when the target and source of the analogy are from different domains. Work has been done to further formalize and potentially quantify distance of analogy using computational text analysis and structuring of design databases [36], but the field on the whole has not yet defined distance of analogy in a more formal way. 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 [37,38].

Design-by-analogy has great potential to produce innovative design. Previous research has shown usage of analogy can mitigate the effects of design fixation [39]. Theoretically, a robust design-by-analogy methodology would enable designers to identify nonobvious 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.

Researchers have worked to facilitate design-by-analogy with a few different strategies, including creating design processes/methodologies, creating ways to translate from one domain of expertise to another, and creating ways to retrieve useful information for analogical stimuli. On the process front, researchers have used problem reformulations, function structures, constraint analysis, and other strategies to achieve design-by-analogy. For example, 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 [40,41]. 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 [42,43]. 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 [44]. Qian and Gero created an exploration medium for between-domain analogies using function-behavior-structure design prototypes [45]. FunSION, a computational tool developed by Liu et al., takes qualitative functional input and output requirements, and generates physical embodiments of design solutions [46,47].

For translating between domains to facilitate analogical transfer of knowledge, Hacco and Shu also developed structured approaches utilizing biomimetic principles for generating concepts, which provides a systematic process for identifying analogous concepts [48]. They also use a functional semantic representation, in which keywords are derived that relates the function to the biological processes. A search is then preformed using standard biological processes from biology textbooks as the reference database. Charlton and Wallace created a web-based tool for finding pre-existing engineering components for reuse in nonstandard applications in new designs to reduce manufacturing costs [49]. Nagel et al. created an engineering to biology thesaurus to reduce the expertise barriers to biologically inspired design [50].

For retrieval of analogical stimuli or potential analogical sources for transfer, 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 [51,52]. Yang et al. worked to create thesauri using information retrieval from informal design documentation for reuse in the design process [53,54], in addition to creating the DedalAI system to automatically index design concepts in electronic notebooks for retrieval and reuse [55]. 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 [56]. Linsey et al. [57–59], Seger et al., [60] and Segers and De Vries [61], and Verhaegen et al. [62] develop approaches to analogical retrieval and reasoning through linguistic (semantic word) associations, problem rerepresentation, and mappings.

Linguistics research has shown that verbs are inherently relational by nature and impose fewer psychological constraints compared to nouns [63]. 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 verbal representation is entirely conceptual, relational, and solution independent.

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 [64]. Additionally, it is acknowledged that the abstract basis functions were derived from a functional model that was created with both functions and flows, and thus the functions by necessity have some latent albeit much reduced dependency/relationship to the flows and solution domain. The representation scheme will use lower level functional (tertiary and correspondents) to specify the design problem. 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 [65]. 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 nonobvious and having not previously existed in the public domain [65].

Another valuable feature of the patent database for design information retrieval is the patent classification structure of the U.S. Patent and Trademark Office (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 [65]. 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 [66,67], 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. The theory of inventive problem solving (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 [68]. Zhang et al. have used the functional basis in combination with TRIZ to create an axiomatic conceptual design model [69]. 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 [70]. To identify relevant candidates for TRIZ automatically, Souli et al. developed a method using linguistic markers [71,72].

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 [36,73,74]. Szykman et al. have built design repositories to share and reuse elements of designs in the development of large scale or complex engineering systems [75]. Koch et al. developed PatViz to allow for visual exploration of queries and complex patent searches using many kinds of patent data, in which the graph views are created by the user [76]. 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 [77]. Chakrabarti et al. used a topic model to analyze patent data, leading to a taxonomy or hierarchical structure [78].

While the U.S. patent database is a ripe repository to support design-by-analogy and is organized by a helpful classification system through the USPTO, the size and complexity make it very difficult to access analogically inspiring information. Attempts to aid in the search and use of the patent database include theories like TRIZ and their resulting tools [68,69,79–89], along with many more research driven tools and methods [40,62,90–92]. 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 Sec. 2, the underlying mathematical formalism to represent functionality is reviewed, the function VSM. 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 Fig. 1. 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 Fig. 1, 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. Figure 2 depicts further detail of the parts involved in this initial step of processing the patent documents. The VSM of information retrieval is used as the basis of the analogy search method developed in this work [93]. 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 [94,95]. 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 [93]. The two common weighting factors are the term frequency ( $tf$ ), which is the frequency of occurrence within a specific document, and the

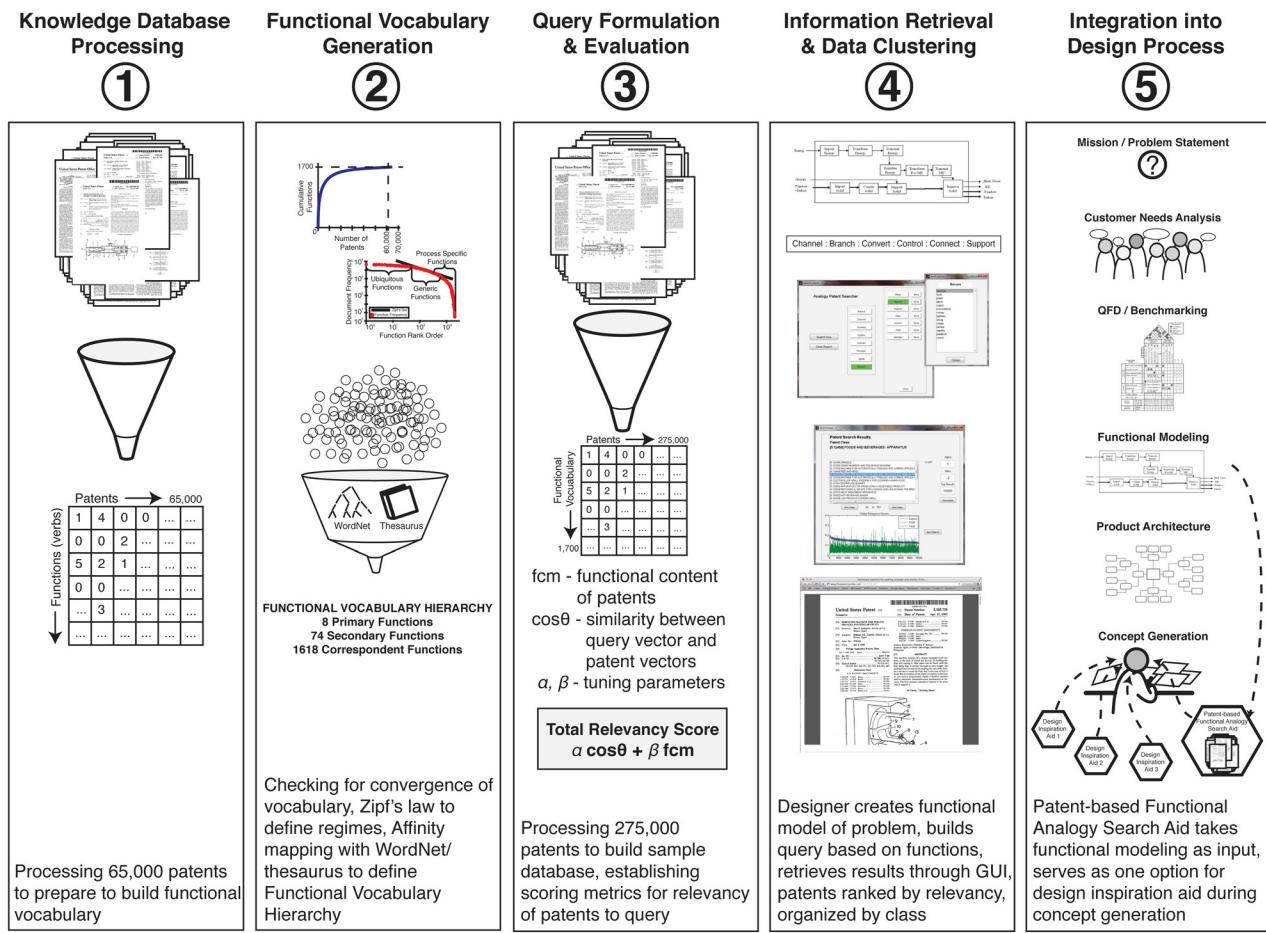


Fig. 1 Overview of the functional analogy search development

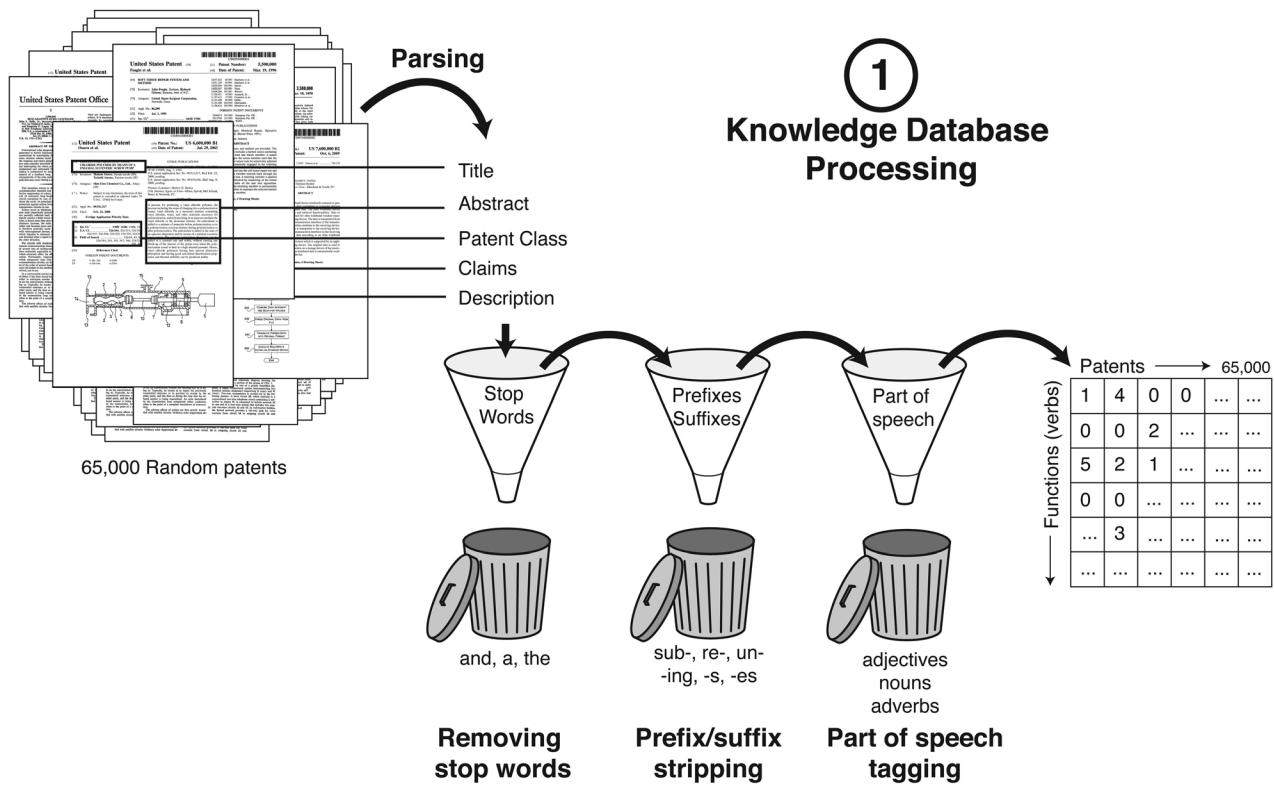


Fig. 2 Knowledge database processing involves parsing, tokenizing, and stemming textual content of 65,000 random patents

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 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 [96]. 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 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 [97].

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]. Conflicting reports of the performance improvement relative to the standard VSM query are reported in the literature. Dumais [97] reports an average performance increase of 5%, and Moldovan et al., [96] found only a 5% improvement over VSM for application of LSI specifically to patent searches. 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, straightforward 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

**Table 1 Examples from the expanded functional basis vocabulary for the secondary functions of *divide* and *import***

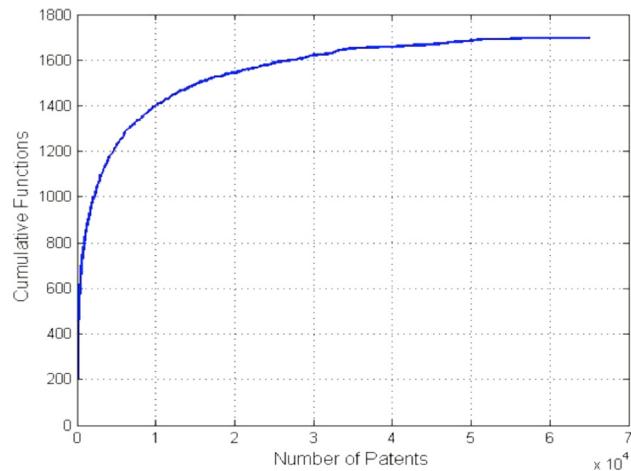
Primary	Secondary	Correspondents
Branch	Divide	Section Divide Segment Branch Sort Partition Tab Miss Diverge Fractionate Cube Segregate Dissociate Graduate Quantize Parse Allot Buck Dissect Dismantle Shred Interdigitate Packetize Compartmentalize Part Separate Digitalize Modularize Butcher Sectionalize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

stemming algorithm is applied to terms to strip suffixes, e.g., -ing, -s, -es [98]. 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 [98]. 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,” and “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.

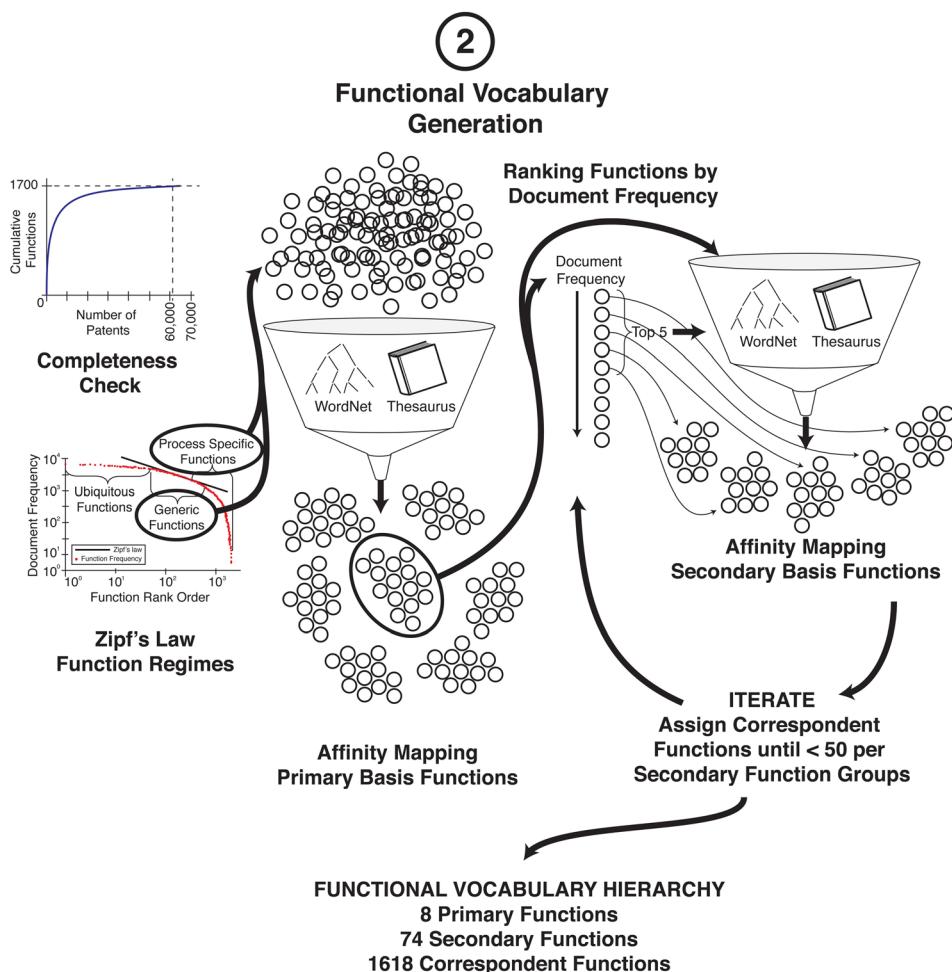
**2.2 Functional Vocabulary Generation.** A primary goal of this research is to identify and extract a complete set of functions covering the entirety of the patent database. Figure 3 depicts the process involved in this second step of the development of the search methodology. Completeness of the function 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 1700 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 Fig. 4, 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 1700 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 Fig. 5 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

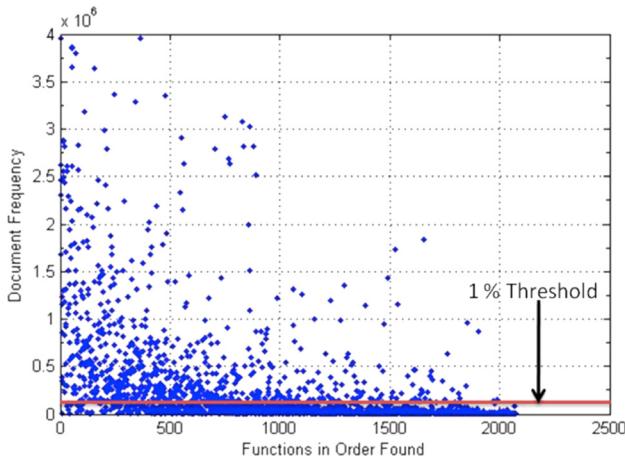


**Fig. 4 Cumulative functions versus number of patents indexed with horizontal asymptote at ~1700 functions and 61,000 patents verifying convergence of function vocabulary**

all patents. Statistically, high document frequency terms will be found earlier due to the random sampling. The trend shown in Fig. 5 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,



**Fig. 3 Functional vocabulary generation involves checking for convergence of functional terms, defining function regimes using zipf's law, and using wordnet/thesaurus to perform affinity mapping, defining functional vocabulary hierarchy**



**Fig. 5 Term-document frequency versus order, showing the frequency falls below a 1% threshold (occur in <~45,000 patents)**

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 [11,12,94], 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 Fig. 5, using log-log axes. A Zipf distribution was fit through the data for comparative purposes, as shown in Fig. 6, quantifying the resolving power of different terms.

Examining Fig. 6, 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

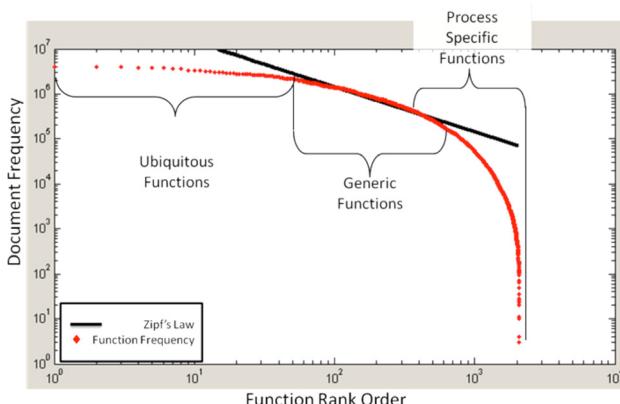
theory. These functions can be considered to lie above the upper cut-off, chosen to be all terms that occur in more than 50% of patents. Examples of these functions are *provide*, *use*, etc. The ubiquitous functions, which account for 50 of the 1700 terms, are to be removed from the final function vocabulary 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 1700 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 1700 word functional vocabulary, modeled after the functional basis [7,8]. The affinity diagram technique is used to group like-terms together into subgroups 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 subfunctions. The function subgroups were split or merged accordingly to attain consistent numbers of functions in each subgroup. 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.
- (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.

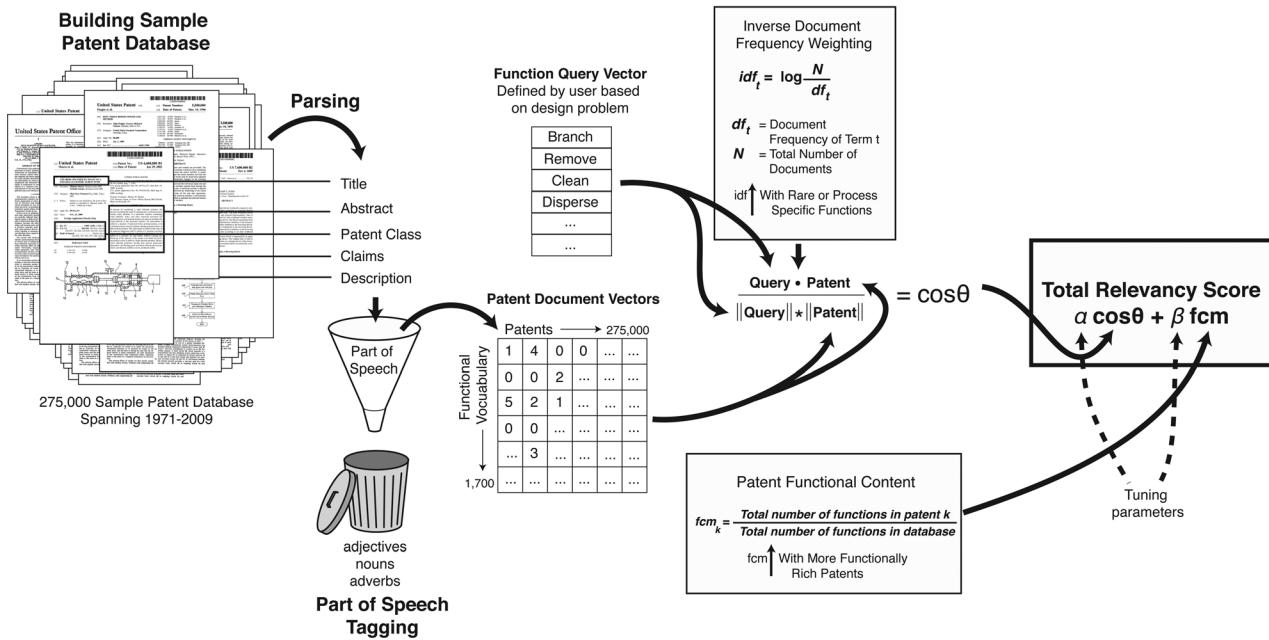
The resulting structure of the expanded functional basis vocabulary is 1700 unique functions organized into 74 groups of secondary functions. The secondary functions and associated correspondents are mapped into the eight (8) primary functions. Table 1 illustrates the hierarchical structure for two of 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 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



**Fig. 6 Function vocabulary document frequency versus rank order comparison with zipf's power law distribution**

### 3 Query Formulation and Evaluation



**Fig. 7** Query formulation and evaluation involves creating a sample patent database of 275,000 patents, defining how to build query vectors for chosen primary and secondary functions, and establishing a relevancy scoring for any patent in the database to a given functional query

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 2,75,000 patents mapped into document vectors, resulting in an approximately  $275,000 \times 1700$  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 Fig. 1) of the research was to formulate the means to query the database of patents and functions. The detailed process for this third step in the development of the methodology is depicted in Fig. 7. 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 [99], 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 metric and the *patent functional content* metric. This is summarized in Table 2.

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

**Table 2** Metrics for calculating similarity between the document and query vectors

– Query-Patent cosine similarity	$\cos \theta = \frac{\text{Query} \cdot \text{Patent}}{\ \text{Query}\  \cdot \ \text{Patent}\ }$
– Patent functional content	$FCM = \frac{\sum \text{Patent}_{\text{terms}(t)}}{\text{NumTerms}}$
– Total relevancy criteria	$\text{Score} = \alpha \cdot \cos \theta + \beta \cdot FCM$

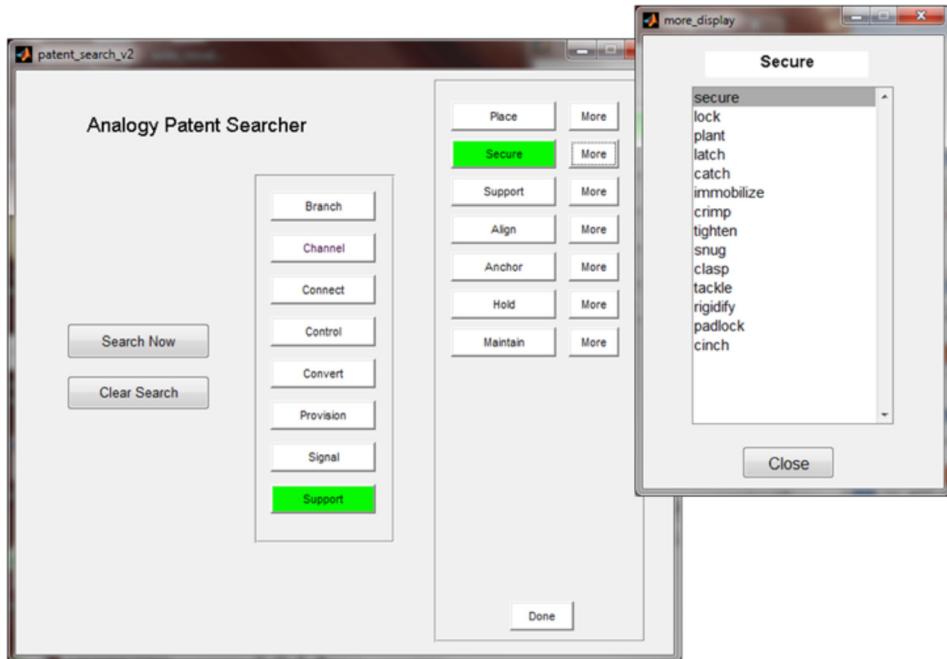


Fig. 8 Query generator user interface

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 graphical user interface builds the query using the expanded functional basis vocabulary hierarchical structure. First, as shown in Fig. 8, 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 Sec. 3.

**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 Fig. 1. This step is depicted in greater detail in Fig. 9. This is implemented in a *search result viewer*, shown in Fig. 10. 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

$$\cos_{\text{similarity}} = q^T \cdot d \quad (3)$$

where  $\cos_{\text{similarity}}$  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  $\cos_{\text{vector}}$  and the functional content metric vector, weighted by the user-defined  $\alpha$  and  $\beta$  coefficients, respectively.

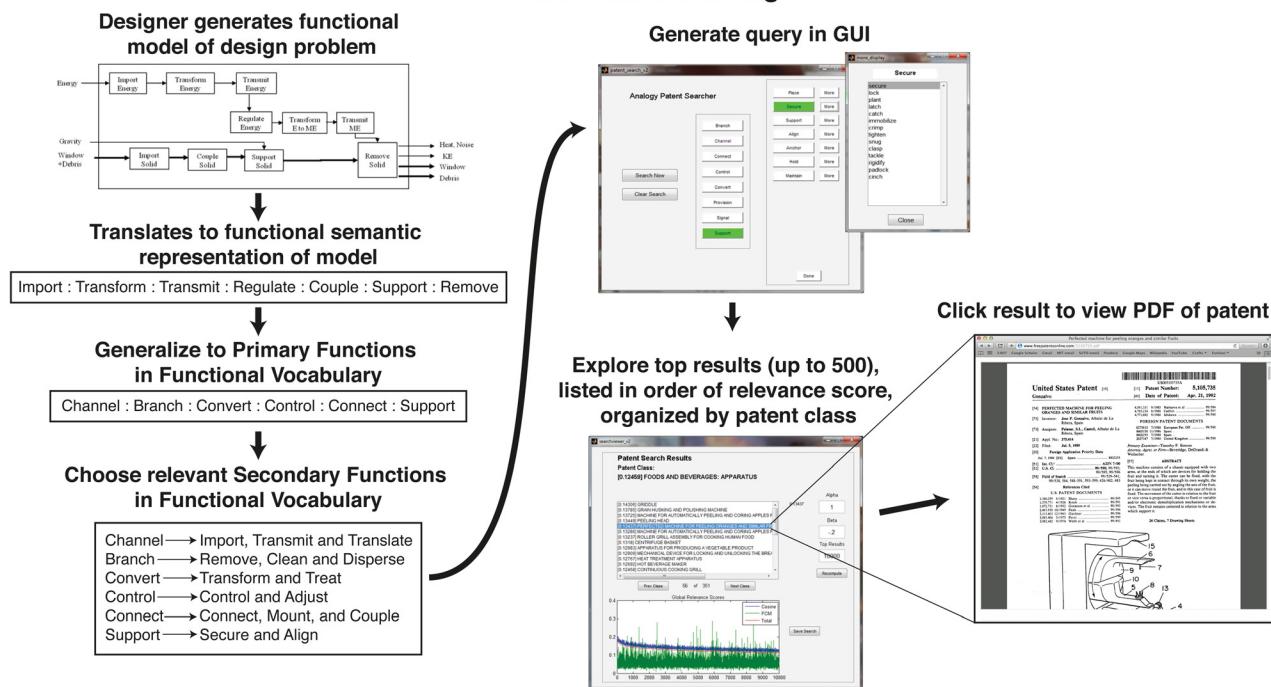
The top  $n$  results as specified by the user are retrieved, sorted by total relevancy score and clustered by primary patent classification. As shown in Fig. 11, the similarity scores for the individual patents are clearly indicated in the first column of the results list. The average relevancy score 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 portable document format (PDF) version of the selected patent, by making calls to online patent databases such as<sup>2</sup>, 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  $\beta$  is varied from 1 to  $-1$  keeping  $\alpha = 1$ , the search results provided more functionally relevant results for negative values of the fcm coefficient  $\beta$ . 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  $\beta$  skew the results toward long patents since, 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  $\alpha$  and  $\beta$  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 Fig. 1) is to make use of the resulting patents presented. The steps described in Secs. 2.1–2.4 are combined into a structured methodology for identifying analogous patents. With the concept generation process, the analogy search methodology

<sup>2</sup>Freepatentsonline.com



**Fig. 9** Information retrieval and data clustering involves entering desired primary and secondary functions into query generator, exploring top 500 patents in search result viewer clustered by uspto patent class, viewed in PDF form with online patent database website

is used as a supplemental technique to more traditional concept generation methods, such as brainstorming, brainsketching, and the CSketch/6-3-5 method [8,100–102]. Figure 11 depicts how the analogous patent search presented in this paper might fit into a product design workflow. 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 Case Study Evaluation of Conceptual Design Phase Efficacy

Two case studies utilized to evaluate the methodology presented in this paper is the design of an automated window washing device and the design of a guitar pick up winder. For the first, 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 such as the power source and cleaning method. The blackbox 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 Fig. 12. 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.

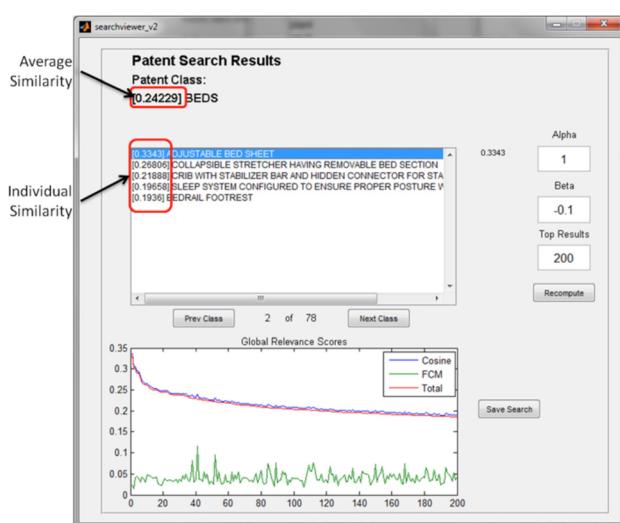
The functional semantic representation of the simplified model becomes

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



**Fig. 10** Search result viewer showing average total relevancy score for patent class and individual total relevancy score for identified patents

## 5 Integration into Design Process

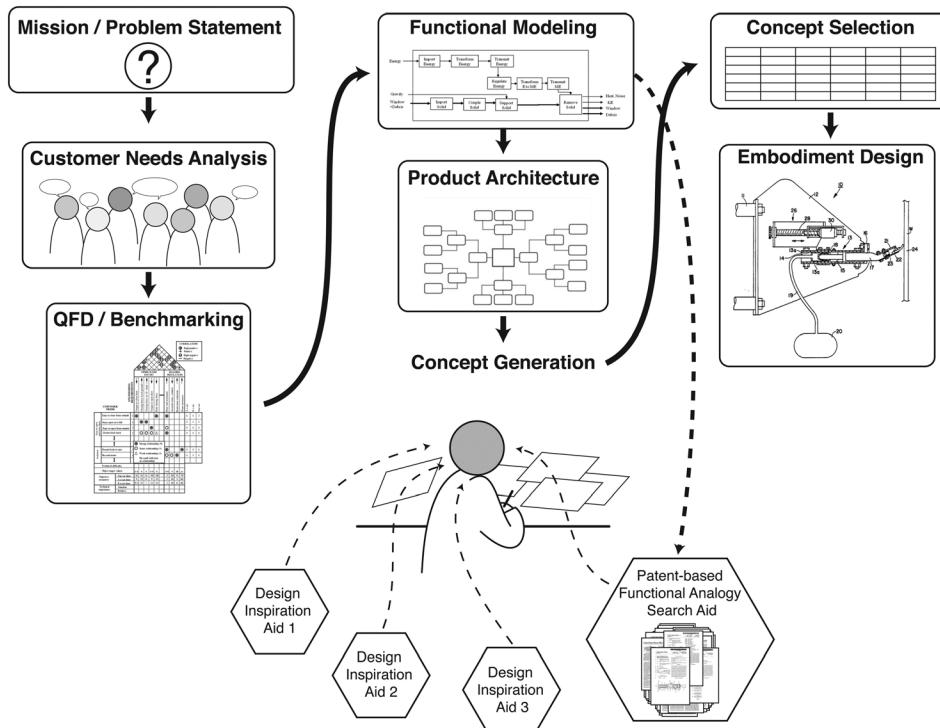


Fig. 11 Integration into design process involves input from user generated functional modeling of design problem into patent analogy search, use as one of many possible design inspiration methods/aids during concept generation

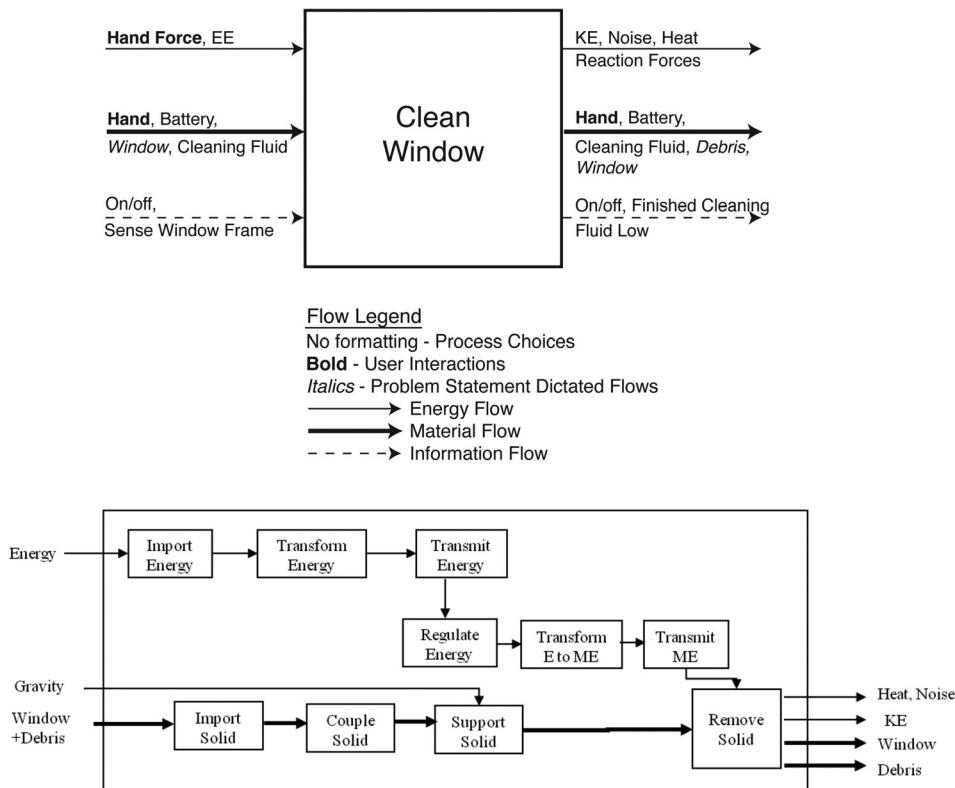


Fig. 12 Black box functional model (top) and simplified functional model (bottom) of core functionality for an automated window washer

Table 3 Combined search results for the automated window washer

Patent title	Patent nos.
[0.19507] Methods for cleaning materials	7556654
[0.15162] Autonomous floor-cleaning robot	6883201
[0.14184] Swimming pool vacuum cleaner with rotary brush	5044034
[0.13631] Device for cleaning a window glass	5086533
[0.1916] Powered cleaner/polisher	7565712
[0.13631] Method and apparatus for cryogenic removal of solid materials	5025632
[0.15685] Washing Device for cleaning a cylinder of a printing machine	5035178
[0.13624] Method of polishing a semiconductor wafer	7559825
[0.18867] Vehicle washing apparatus	5077859
[0.15076] Apparatus for supporting a direct drive drilling unit	5038871
[0.15398] Construction elevator assembly	5033586
[0.17476] Method and system for maintaining equal and continuous flows of liquid to and from intermittently operating apparatus	3589389

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. Table 3 summarizes the relevant patents compiled from the search results for the queries listed above.

The fourth patent identified for the window cleaning device (Patent No. 5,086,533) is a very near-field analogy to the proposed design problem. The device shown in Fig. 13(a) utilizes a squeegee mechanism with a fluid application system to automatically clean windows. A second cleaning device, shown in Fig. 13(b), is used for automatically cleaning floors.

The second patent identified is a floor cleaning robot (Patent No. 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 eighth patent identified, a wafer polishing patent (Patent No. 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 No. 5,033,586 for a transportable construction elevator, shown in Fig. 13(c), using a pulley mechanism.

Finally, entirely novel methods of cleaning surfaces are identified using the patent-based functional analogy search methodology. The sixth patent identified, Patent No. 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. 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.

The second case study was chosen to illustrate that the computational method presented in this paper could be utilized to reproduce, if not improve on, the solutions for the classic design-by-

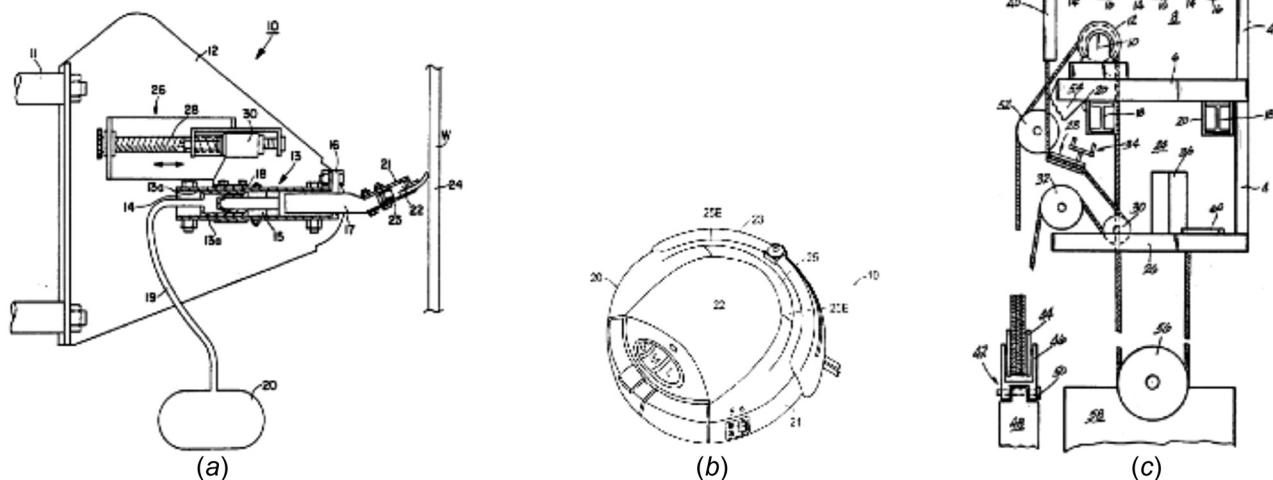


Fig. 13 Patent analogy search results. (a) Automated window cleaning device, an example of a near-field analogy (Patent No. 5,086,533), (b) automated floor cleaning device, an example of a far-field analogy (Patent No. 6,883,201), (c) transportable elevator system for vertically traversing buildings under construction (Patent No. 5,033,586).

**Table 4 Results of similarity calculation for the pickup winder [103]**

Product	Similarity index, $\lambda$
Pickup winder	1.0
Fruit and vegetable peeler	0.78
Electric can opener	0.74
Electric sander	0.73
Fishing reel	0.72
Cheese grater	0.69

analogy problem of the guitar pickup winder. McAdams and Wood [103] compare the functionality of the pickup winder, which is used to manufacture electromagnetic coils for electric guitars, to a database of 68 products. The five most functionally similar products are shown in Table 4, where  $\lambda$  is the normalized similarity index.

In Fig. 14, a simplified functional model of the pickup winder is given which includes the top six functions as determined from the weightings of the corresponding customer needs. The generic top six functions are: *import*, *secure*, *position*, *regulate*, *guide*, and *allow rotational degrees of freedom (DOF)* as shown in the solid boxes. In the expanded patent functional basis, the functional semantic representation maps as follows:

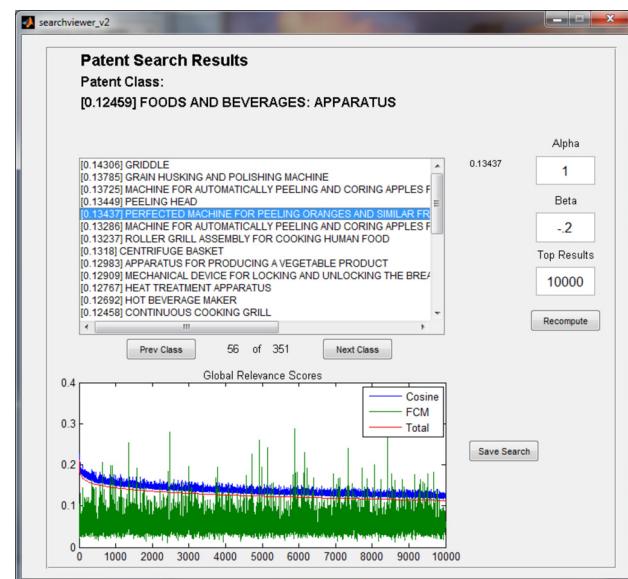
- Import: Channel → Import
- Secure: Support → Secure
- Position: Support → Place
- Regulate: Control → Control
- Guide: Channel → Direct
- Allow rotational DOF: Channel → Rotate

After the mapping from the original functional model to the expanded basis is established, the search for analogous patents is performed utilizing relevance ranking weights of  $\alpha=1$  and  $\beta=-0.2$  and the top 10,000 results are reviewed.

The first phase of the case study was to determine whether the patent search could extract the analogous products from Table 4. Figure 15 shows the example of the search result for the fruit peeler. Considering the relative sparseness of patents included in the prototype database ( $\sim 6\%$  of electronically available patents), the search results are very successful with the search coverage including three of the five top analogous devices.

Figure 16 depicts sample illustrations for the three analogous devices retrieved: a fruit peeler, a fishing reel with disengageable spool, and a belt sander.

It must be noted that a large pool of search results is required to identify the three analogous patents. The spinning reel and belt sander occurred within the top 1000 search results, but the fruit peeler is not retrieved until a group of the top 10,000 results are extracted. The large number of patents required to extract the fruit peeler analogy is caused by a highly populated query vector

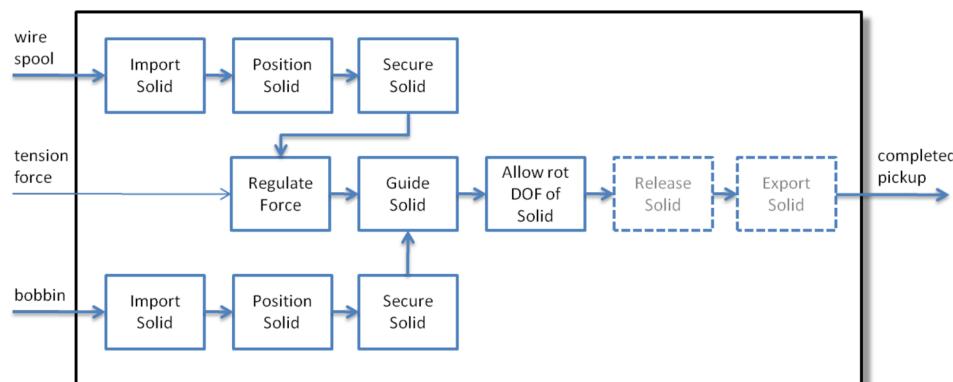


**Fig. 15 Search results for pickup winder generic functions showing fruit peeler analogy**

resulting from multiple Secondary functions used in the query generation. With each Secondary function mapping to an average of 20 correspondents, the total number of terms in the query vector is approximately 120 terms and leads to poor resolution with respect to the cosine similarity metric. One of the significant insights gained through this case study is multiple searches on individual functions improves discrimination among the search results. Despite the relevancy resolution issues encountered, an additional patent was found that, if implemented into the pickup winder design, would provide a novel means of controlling the wire tension.

In Fig. 17, a tension control mechanism for a musical instrument excites the wire with a known frequency and measures the response. The tuning device automatically adjusts the tension to bring the frequency response to within the specified range. A device using piezoelectric actuators to control the tension, actuate the wire and sense the frequency response can be designed into an advanced version of the pickup winder if constant tension is critical for enhanced pickup performance or process control consistency.

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 Refs. [104,105].



**Fig. 14 Simplified functional model of a guitar pickup winder**

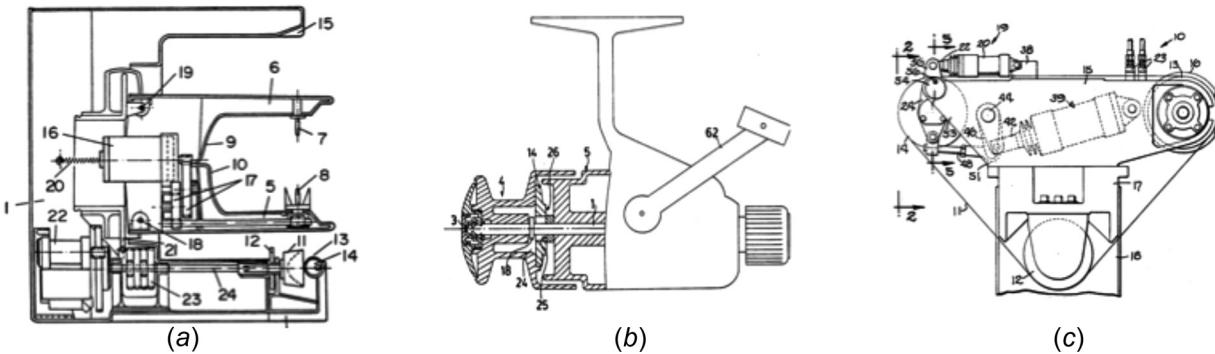


Fig. 16 Illustrations from analogous patents for the pickup winder. (a) U.S. Patent No. 5,105,735: perfected machine for peeling oranges and similar fruits, (b) U.S. Patent No. 5,121,888: spinning reel with a spool disengageable from a rear-mounted drag, (c) U.S. Patent No. 3,577,684: abrading machine with steering roll and tensioned abrading belt.

## 4 Limitations and Future Work

The computational methodology presented in this paper has been shown to be effective in two case studies: identifying functional analogies from the patent database and increasing the novelty of concepts generated during ideation. Significant further experimental validation is required to ensure that the methodology is indeed effective for multiple types of design problems in many different contexts. Although the initial results are promising, continued improvements to the search engine design could further enhance the tool's efficacy. The first improvement proposed is simply increasing the patent database coverage. The computational capacity of the current prototype could be refined into a more user-friendly tool.

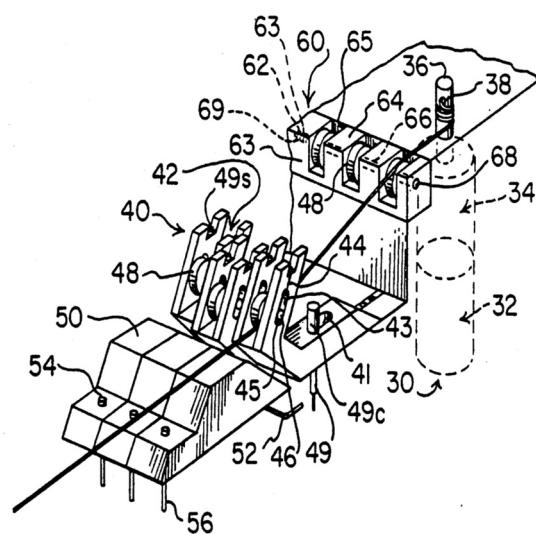
Further research is needed to optimize the *total relevancy score* metric. Including patent length normalization in the patent functional content metric could minimize the bias resulting from longer patents including a broader range of function terms. Additionally, a rigorous experimental study to determine the optimal relevancy score coefficient weights must be conducted to verify the results from the parametric process utilized in this research.

Additional extensions to the search engine to be investigated are the inclusion of customer needs utilizing system attribute terms as adjectives. The attribute terms would be implemented as context limiters used to augment the similarity metric. Some examples of attribute term adjectives are *quickly, cheap, light*, etc. 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. Finally, a potential further expansion of the work is into the exploration of the knowledge and learning that occurs when analogical transfer occurs between a source and target field, as is facilitated with this methodology.

## 5 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 1700 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. Two case studies were conducted to evaluate the methodology. Despite the limited patent coverage in the prototype patent database, the search process extracted three of the five top analogous products for the pickup winder case study, and showed promisingly useful results for the window washer case study as well. Key insights gained from the case studies were generating large query vectors by searching over multiple primary and secondary functions detrimentally reduces the total relevancy score metric resolution, requiring the user to search many more patents to extract the desired patent analogies. Going forward, the search approach is best used by performing multiple searches over fewer functions. The computational methodology presented in this paper shows promise as the foundation of an automated function-based design-by-analogy inspiration tool.



**Fig. 17** Wire tension control analogy solution, U.S. Patent No. 5,038,657

## Acknowledgment

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 Nos. CMMI-0855326, CMMI-0855510, and CMMI-0855293, and the SUTD-MIT international Design Centre (IDC) (<http://idc.sutd.edu.sg/>).

## References

[1] Salton, G., 1971, *The SMART Retrieval System—Experiments in Automatic Document Retrieval*, Prentice Hall, Englewood Cliffs, NJ.

[2] Salton, G., Wong, A., and Yang, C. S., 1975, "A Vector Space Model for Automatic Indexing," *Commun. ACM*, **18**(11), pp. 613–620.

[3] Schmid, H., 1994, "Probabilistic Part-of-Speech Tagging Using Decision Tree," International Conference on New Methods in Language Processing, Manchester, UK, Vol. 12, pp. 44–49.

[4] Kurfman, M., Rajan, J., Stone, R., and Wood, K., 2001, "Functional Modeling Experimental Studies," *ASME* Paper No. DETC2001-21709.

[5] Kurfman, M., Rajan, J., Stone, R., and Wood, K., 2003, "Experimental Studies Assessing the Repeatability of a Functional Modeling Derivation Method," *ASME J. Mech. Des.*, **125**(4), pp. 682–693.

[6] Stone, R., and Wood, K. L., 2000, "Development of a Functional Basis for Design," *ASME J. Mech. Des.*, **122**(4), pp. 359–370.

[7] Hirtz, J., Stone, R. B., McAdams, D. A., Szykman, S., and Wood, K. L., 2002, "A Functional Basis for Engineering Design: Reconciling and Evolving Previous Efforts," *Res. Eng. Des.*, **13**, pp. 65–82.

[8] Otto, K., and Wood, K., 2001, *Product Design Techniques in Reverse Engineering and New Product Development*, Prentice Hall, Upper Saddle River, NJ.

[9] Manning, C. D., Raghavan, P., and Schütz, H., 2009, *An Introduction to Information Retrieval*, Cambridge University, Cambridge, UK.

[10] Salton, G., and Waldstein, R. K., 1978, "Term Relevance Weights in On-Line Information Retrieval," *Inform. Process. Manage.*, **14**(1), pp. 29–35.

[11] Wyllys, E. E., 1981, "Empirical and Theoretical Bases of Zipf's Law," *Libr. Trends*, **30**(1), pp. 53–64.

[12] Zipf, G. K., 1949, *Human Behavior and the Principle of Least Effort*, Addison-Wesley, Oxford, UK.

[13] Fellbaum, C., 1998, *WordNet: An Electronic Lexical Database*, MIT, Cambridge, MA.

[14] Miller, G. A., 1995, "WordNet: A Lexical Database for English," *Commun. ACM*, **38**(11), pp. 39–41.

[15] Pahl, G., and Beitz, W., 1996, *Engineering Design: A Systematic Approach*, 2nd ed., Springer-Verlag, London, UK.

[16] Chan, J., Fu, K., Schunn, C., Cagan, J., Wood, K., and Kotovsky, K., 2011, "On the Benefits and Pitfalls of Analogies for Innovative Design: Ideation Performance Based on Analogical Distance, Commonness, and Modality of Examples," *ASME J. Mech. Des.*, **133**(8), p. 081004.

[17] Linsey, J., Murphy, J., Markman, A., Wood, K. L., and Kortoglu, T., 2006, "Representing Analogies: Increasing the Probability of Innovation," *ASME* Paper No. DETC2006-99383.

[18] Chakrabarti, A., and Bligh, T. P., 2001, "A Scheme for Functional Reasoning in Conceptual Design," *Des. Stud.*, **22**(6), pp. 493–517.

[19] Cheong, H., Chiu, I., Shu, L. H., Stone, R., and McAdams, D., 2011, "Biologically Meaningful Keywords for Functional Terms of the Functional Basis," *ASME J. Mech. Des.*, **133**(2), p. 021007.

[20] Shu, L. H., Ueda, K., Chiu, I., and Cheong, H., 2011, "Biologically Inspired Design," *CIRP Ann.*, **60**(2), pp. 673–693.

[21] Stone, R., and Wood, K., 2000, "A Heuristic Method for Identifying Modules for Product Architectures," *Des. Stud.*, **21**(1), pp. 5–31.

[22] Salonen, M., Holltta-Otto, K., and Otto, K., 2008, "Effecting Product Reliability and Life Cycle Costs With Early Design Phase Product Architecture Decisions," *Int. J. Prod. Dev.*, **5**(1/2), pp. 109–124.

[23] Bohm, M., and Stone, R., 2004, "Product Design Support: Exploring a Design Repository System," *ASME* Paper No. IMECE2004-61746.

[24] Bohm, M. R., Vucovich, J. P., and Stone, R. B., 2005, "Capturing Creativity: Using a Design Repository to Drive Concept Innovation," *ASME* Paper No. DETC2005-85105.

[25] Bohm, M., and Stone, R., 2004, "Representing Functionality to Support Reuse: Conceptual and Supporting Functions," *ASME* Paper No. DETC2004-57693.

[26] Szykman, S., Sriram, R. D., Bochenek, C., and Racz, J., 1999, "The NIST Design Repository Project," *Advances in Soft Computing—Engineering Design and Manufacturing*, R. Roy, T. Furuhashi, and P. K. Chawdhry, eds., Springer-Verlag, London.

[27] Szykman, S., Sriram, R. D., Bochenek, C., and Senfaute, J., 2000, "Design Repositories: Next-Generation Engineering Design Databases," *IEEE Intell. Syst. Appl.*, **15**(3), pp. 48–55.

[28] Bryant, C., Stone, R., McAdams, D., Kurtoglu, T., and Campbell, M., 2005, "Concept Generation From the Functional Basis of Design," 15th International Conference on Engineering Design: Engineering Design and the Global Economy, A. Samuel and W. Lewis (eds.), Barton, A.C.T. Engineers, Australia, pp. 1702–1715.

[29] Bryant, C., Stone, R., McAdams, D., Kurtoglu, T., and Campbell, M., 2005, "A Computational Technique for Concept Generation," *ASME* Paper No. DETC2005-85323.

[30] Terpenny, J., and Mathew, D., 2004, "Modeling Environment for Function-Based Conceptual Design," *ASME* Paper No. DETC2004-57492.

[31] Potter, S., Culley, S. J., Darlington, M. J., and Chawdhry, P. K., 2003, "Automatic Conceptual Design Using Experience-Derived Heuristics," *Res. Eng. Des.*, **14**(3), pp. 131–144.

[32] Chiu, M., 2003, "Design Moves in Situated Design With Case-Based Reasoning," *Des. Stud.*, **24**(1), pp. 1–25.

[33] Gentner, D., 1983, "Structure Mapping—A Theoretical Framework for Analogy," *Cognit. Sci.*, **7**, pp. 155–170.

[34] Falkenhainer, B. F., Forbus, K. D., and Gentner, D., 1989, "The Structure Mapping Engine: Algorithm and Examples," *Artif. Intell. Eng. Des. Anal. Manuf.*, **41**(1), pp. 1–63.

[35] Gentner, D., and Markman, A. B., 1997, "Structure Mapping in Analogy and Similarity," *Am. Psychol.*, **52**(1), pp. 45–56.

[36] Fu, K., Chan, J., Cagan, J., Kotovsky, K., Schunn, C., and Wood, K., 2013, "The Meaning of "Near" and "Far": The Impact of Structuring Design Databases and the Effect of Distance of Analogy on Design Output," *ASME J. Mech. Des.*, **135**(2), p. 021007.

[37] Kryssanov, V. V., Tamaki, H., and Kitamura, S., 2001, "Understanding Design Fundamentals: How Synthesis and Analysis Drive Creativity, Resulting in Emergence," *Artif. Intell. Eng.*, **15**(4), pp. 329–342.

[38] Goldschmidt, G., and Weil, M., 1998, "Contents and Structure in Design Reasoning," *Des. Issues*, **14**(3), pp. 85–100.

[39] Linsey, J., Tseng, I., Fu, K., Cagan, J., Wood, K., and Schunn, C., 2010, "A Study of Design Fixation, Its Mitigation and Perception in Engineering Design Faculty," *ASME J. Mech. Des.*, **132**(4), p. 1041003.

[40] Goel, A., Bhatta, S., and Stroulia, E., 1997, "Kritik: An Early Case-Based Design System," *Issues and Applications of Case-Based Reasoning in Design*, M. Maher and P. Pu, eds., Erlbaum, Mahwah, NJ, pp. 87–132.

[41] Goel, A., and Chandrasekaran, B., 1989, "Functional Representation of Designs and Redesign Problem Solving," Proceedings of the Eleventh International Joint Conference on Artificial Intelligence, pp. 1388–1394.

[42] Bhatta, S., Goel, A., and Prabhakar, S., 1994, "Innovation in Analogical Design: A Model-Based Approach," *Artificial Intelligence in Design'94*, Springer Netherlands, Heidelberg, Germany, pp. 57–74.

[43] Goel, A., and Bhatta, S., 2004, "Use of Design Patterns in Analogy-Based Design," *Adv. Eng. Inf.*, **18**(2), pp. 85–94.

[44] Navinchandra, D., Sycara, K. P., and Narasimhan, S., 1991, "Behavioral Synthesis in CADET, a Case-Based Design Tool," Seventh IEEE Conference on Artificial Intelligence Applications, Vol. 1, pp. 217–221.

[45] Qian, L., and Gero, J. S., 1992, "A Design Support System Using Analogy," *Artificial Intelligence in Design*, J. S. Gero and F. Sudweeks, eds., Springer, The Netherlands, Heidelberg, Germany, pp. 795–813.

[46] Liu, Y.-C., Chakrabarti, A., and Bligh, T. P., 2000, *Artificial Intelligence in Design'00*, J. S. Gero (ed.), Springer Netherlands, Heidelberg, Germany, pp. 499–519.

[47] Liu, Y.-C., Bligh, T. P., and Chakrabarti, A., 2003, "Towards an 'Ideal' Approach for Concept Generation," *Des. Stud.*, **24**(4), pp. 341–355.

[48] Hacco, E., and Shu, L. H., 2002, "Biomimetic Concept Generation Applied to Design for Remanufacture," *ASME* Paper No. DETC2002/DFM-34177.

[49] Charlton, C. T., and Wallace, K. M., 2000, "A Web Broker for Component Retrieval in Mechanical Engineering," *Des. Stud.*, **21**(2), pp. 167–186.

[50] Nagel, J. K., Stone, R., and McAdams, D., 2010, "An Engineering-to-Biology Thesaurus for Engineering Design," *ASME* Paper No. DETC2010-28233.

[51] Chakrabarti, A., Sarkar, P., Leelavathama, B., and Nataraju, B. S., 2005, "A Functional Representation for Aiding in Biomimetic and Artificial Inspiration of New Ideas," *Artif. Intell. Eng. Des. Anal. Manuf.*, **19**(2), pp. 113–132.

[52] Chakrabarti, A., 2009, "Design Creativity Research," *Product Research*, N. R. Srinivas Raghavan and J. A. Cafeo, eds., Springer Netherlands, Heidelberg, Germany, pp. 17–39.

[53] Yang, M. C., Wood, W. H., and Cutkosky, M. R., 2005, "Design Information Retrieval: A Thesauri-Based Approach for Reuse of Informal Design Information," *Eng. Comput.*, **21**(2), pp. 177–192.

[54] Wood, W. H., Yang, M. C., Cutkosky, M. R., and Agogino, A. M., 1998, "Design Information Retrieval: Improving Access to the Informal Side of Design," ASME Design Engineering Technical Conference, Atlanta, GA.

[55] Yang, M. C., and Cutkosky, M. R., 1997, "Automated Indexing of Design Concepts for Information Management," International Conference on Engineering Design, Tampere, Schriftenreihe WDK, pp. 191–196.

[56] Ahmed, S., 2005, "An Approach to Assist Designers With Their Queries and Designs," *ASME* Paper No. DETC2006-99705.

[57] Linsey, J., Markman, A. B., and Wood, K. L., 2008, "WordTrees: A Method for Design-by-Analogy," ASEE Annual Conference.

[58] Linsey, J., Markman, A. B., and Wood, K. L., 2012, "Design by Analogy: A Study of the WordTree Method for Problem Re-Representation," *ASME J. Mech. Des.*, **134**(4), p. 041009.

[59] Linsey, J. S., Wood, K. L., and Markman, A. B., 2008, "Increasing Innovation: Presentation and Evaluation of the WordTree Design-by-Analogy Method," *ASME* Paper No. DETC2008-49317.

[60] Segers, N. M., De Vries, B., and Achten, H. H., 2005, "Do Word Graphs Stimulate Design?," *Des. Stud.*, **26**(6), pp. 625–647.

[61] Segers, N., and De Vries, B., 2003, "The Idea Space System: Words as Handles to a Comprehensive Data Structure," 10th International Conference on Computer Aided Architectural Design Futures, Digital Design—Research and Practice, Kluwer Academic Publishers, Dordrecht.

[62] Verhaegen, P., D'hondt, J., Vandevenne, D., Dewulf, S., and Duflou, J. R., 2011, "Identifying Candidates for Design-by-Analogy," *Comput. Ind.*, **62**(4), pp. 446–459.

[63] Gentner, D., 1981, "Some Interesting Differences Between Verbs and Nouns," *Cognit. Brain Theory*, **4**(2), pp. 161–178.

[64] Chiu, I., and Shu, L. H., 2007, "Using Language as Related Stimuli for Concept Generation," *Artif. Intell. Eng. Des. Anal. Manuf.*, **21**(2), pp. 103–121.

[65] Kang, I., Na, S., Kim, J., and Lee, J., 2007, "Cluster-Based Patent Retrieval," *Inform. Process. Manage.*, **43**(5), pp. 1173–1182.

[66] Trippé, A. J., 2003, "Patinformatics: Tasks to Tools," *World Pat. Inform.*, **25**(3), pp. 211–221.

[67] Tseng, Y., Lin, C., and Lin, Y., 2007, "Test Mining Techniques for Patent Analysis," *Inf. Process. Manage.*, **43**(5), pp. 1216–1247.

[68] Altshuller, G. S., and Shapiro, R. B., 1956, "О Психологии изобретательского творчества (On the Psychology of Inventive Creation)," Вопросы Психологии (Psychol. Issues), **6**, pp. 37–39 (in Russian).

[69] Zhang, R., Cha, J., and Lu, Y., 2007, "A Conceptual Design Model Using Axiomatic Design, Functional Basis and TRIZ," *Proceedings of the 2007 IEEE IEEM*, Singapore, pp. 1807–1810.

[70] Cascini, G., and Russo, D., 2007, "Computer-Aided Analysis of Patents and Search for TRIZ Contradictions," *Int. J. Prod. Dev.*, 4(1/2), pp. 52–67.

[71] Souli, A., Cavallucci, D., Rousset, F., and Zanni, C., 2011, "Starting From Patents to Find Inputs to the Problem Graph Model of IDM-TRIZ," *TRIZ Future 2011*, Dublin, Ireland.

[72] Souli, A., and Cavallucci, D., 2012, "Toward an Automatic Extraction of IDM Concepts From Patents," *CIRP Design*, Bangalore, India, Apr. 28–30.

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

[74] Fu, K., Cagan, J., Kotovsky, K., and Wood, K., 2013, "Discovering Structure in Design Databases Through Function and Surface Based Mapping," *ASME J. Mech. Des.*, 135(3), p. 031006.

[75] Szykman, S., Sriram, R. D., Bochenek, C., Racz, J. W., and Senfaute, J., 2000, "Design Repositories: Engineering Design's New Knowledge Base," *IEEE Intell. Syst.*, 15, pp. 48–55.

[76] Koch, S., Bosch, H., Giereth, M., and Ertl, T., 2009, "Iterative Integration of Visual Insights during Patent Search and Analysis," *IEEE Symposium on Visual Analytics Science and Technology*, Atlantic City, NJ, pp. 203–210.

[77] Mukherjee, S., Bhuvan, B., and Kankar, P., 2005, "Information Retrieval and Knowledge Discovery Utilizing a BioMedical Patent Semantic Web," *IEEE Trans. Knowl. Data Eng.*, 17(8), pp. 1099–1110.

[78] Chakrabarti, S., Dom, B., Agrawal, R., and Raghavan, P., 1998, "Scalable Feature Selection, Classification and Signature Generation for Organizing Large Text Databases Into Hierarchical Topic Taxonomies," *VLDB J.*, 7(3), pp. 163–178.

[79] Duran-Novoa, R., Leon-Rovira, N., Aguayo-Tellez, H., and Said, D., 2011, "Inventive Problem Solving Based on Dialectical Negation, Using Evolutionary Algorithms and TRIZ Heuristics," *Comput. Ind.*, 62(4), pp. 437–445.

[80] Hernandez, N. V., Schmidt, L. C., and Okudan, G. E., 2012, "Systematic Idea-Generation Effectiveness Study of TRIZ," *ASME Paper No. DETC2012-70294*, 10,1115/DETC2012-70294.

[81] Hernandez, N. V., Schmidt, L. C., and Okudan, G. E., 2012, "Experimental Assessment of TRIZ Effectiveness in Idea Generation," *ASEE Annual Conference*, San Antonio, TX.

[82] Krasnoslobodtsev, V., and Langevin, R., 2005, "TRIZ Application in Development of Climbing Robots," *First TRIZ Symposium*, Japan.

[83] Liang, Y., Tan, R., and Ma, J., 2008, "Patent Analysis With Text Mining for TRIZ," *IEEE ICMIT*, Bangkok, Thailand, pp. 1147–1151.

[84] Nakagawa, T., 2012, "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.

[85] Nix, A. A., Sherret, B., and Stone, R. B., 2011, "A Function Based Approach to TRIZ," *ASME Paper No. DETC2011-47973*.

[86] Houssin, R., and Coulibaly, A., 2011, "An Approach to Solve Contradiction Problems for Safety Integration in Innovative Design Process," *Comput. Ind.*, 62(4), pp. 398–406.

[87] Mann, D., Dewulf, S., Zlotin, B., and Zusman, A., 2003, *Matrix 2003, Updating the TRIZ Contradiction Matrix*, CREAx, Kurtrijk, Belgium.

[88] CREAx, 2012, "CREAx: Creativity for Innovation," <http://www.creax.com>

[89] Goldfire, I. M., 2012, "Invention Machine Goldfire: Unleashing the Power of Research," Available: <http://inventionmachine.com/products-and-services/innovation-software/goldfire-Research/>

[90] Bhatta, S., and Goel, A., 1996, "From Design Experiences to Generic Mechanisms: Model-Based Learning in Analogical Design," *Artif. Intell. Eng. Des. Anal. Manuf.*, 10(2), pp. 131–136.

[91] Chiu, I., and Shu, L. H., 2005, "Bridging Cross-Domain Terminology for Biomimetic Design," *ASME Paper No. DETC2005-84908*.

[92] Vincent, J. F. V., Bogatyreva, O. A., Bogatyreva, N. R., Bowyer, A., and Pahl, A. K., 2006, "Biomimetics: Its Practice and Theory," *J. R. Soc. Interface*, 3(9), pp. 471–482.

[93] Salton, G., and McGill, M. J., 1986, *Introduction to Modern Information Retrieval*, McGraw-Hill, New York, NY.

[94] van Rijsbergen, C. J., 1979, *Information Retrieval*, Butterworth-Heinemann, Oxford, UK.

[95] Rindflesch, T. C., 1996, "Natural Language Processing," *Ann. Rev. Appl. Linguist.*, 16, pp. 70–85.

[96] Moldovan, A., Bot, R. I., and Wanka, G., 2005, "Latent Semantic Indexing for Patent Documents," *Int. J. Appl. Math. Comput. Sci.*, 15, pp. 551–560.

[97] Dumais, S. T., 1995, "Latent Semantic Indexing (LSI): TREC-3 Report," *Proceedings of TREC*.

[98] Porter, M. F., 1980, "An Algorithm for Suffix Stripping," *Program*, 14(3), pp. 130–137.

[99] Linsey, J., Clauss, E. F., Kurtoglu, T., Murphy, J. T., Wood, K. L., and Markman, A. B., 2011, "An Experimental Study of Group Idea Generation Techniques: Understanding the Roles of Idea Representation and Viewing Methods," *ASME J. Mech. Des.*, 133(3), p. 031008.

[100] Osborn, A., 1957, *Applied Imagination*, Scribner, New York, NY.

[101] Markman, A. B., and Wood, K. L., 2009, *Tools for Innovation: The Science Behind Practical Methods that Drive New Ideas*, Oxford University, New York, NY.

[102] Vangundy, A. B., 1988, *Techniques of Structured Problem Solving*, 2nd ed., Van Nostrand Reinhold Company, NY.

[103] McAdams, D., and Wood, K., 2002, "A Quantitative Similarity Metric for Design by Analogy," *ASME J. Mech. Des.*, 124(2), pp. 173–182.

[104] Murphy, J. T., 2011, "Patent-Based Analogy Search Tool for Innovative Concept Generation," Ph.D. thesis, Department of Mechanical Engineering, The University of Texas, Austin, TX.

[105] Fu, K., Murphy, J., Yang, M., Otto, K., Jensen, D., and Wood, K. L., 2013, "Investigating the Effect of Functionality Level of Analogical Stimulation on Design Outcomes," *Korea-Japan Design Engineering Workshops (DEWS)*, Kitakyushu, Fukuoka, Japan, Nov. 28–30.