Biomimetic design through natural language analysis to facilitate cross-domain information retrieval

IVEY CHIU AND L.H. SHU

Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Ontario, Canada

Abstract

Biomimetic, or biologically inspired, design uses analogous biological phenomena to develop solutions for engineering problems. Several instances of biomimetic design result from personal observations of biological phenomena. However, many engineers' knowledge of biology may be limited, thus reducing the potential of biologically inspired solutions. Our approach to biomimetic design takes advantage of the large amount of biological knowledge already available in books, journals, and so forth, by performing keyword searches on these existing natural-language sources. Because of the ambiguity and imprecision of natural language, challenges inherent to natural language processing were encountered. One challenge of retrieving relevant cross-domain information involves differences in domain vocabularies, or lexicons. A keyword meaningful to biologists may not occur to engineers. For an example problem that involved cleaning, that is, removing dirt, a biochemist suggested the keyword "defend." Defend is not an obvious keyword to most engineers for this problem, nor are the words defend and "clean/remove" directly related within lexical references. However, previous work showed that biological phenomena retrieved by the keyword defend provided useful stimuli and produced successful concepts for the clean/remove problem. In this paper, we describe a method to systematically bridge the disparate biology and engineering domains using natural language analysis. For the clean/remove example, we were able to algorithmically generate several biologically meaningful keywords, including defend, that are not obviously related to the engineering problem. We developed a method to organize and rank the set of biologically meaningful keywords identified, and confirmed that we could achieve similar results for two other examples in encapsulation and microassembly. Although we specifically address cross-domain information retrieval from biology, the bridging process presented in this paper is not limited to biology, and can be used for any other domain given the availability of appropriate domain-specific knowledge sources and references.

Keywords: Biomimetic, Biologically Inspired, Bioanalogous Design; Concept Generation; Domain Lexicons; Information Retrieval

1. INTRODUCTION

Biomimetic design uses biological phenomena as inspiration for solutions to engineering problems. One well-known example of biomimetic design is the development of Velcro after observing that cockleburs attach to clothing and fur.

In the development of synectics, Gordon (1961) observed that biology provided the richest source of direct analogies. Benami and Jin (2002) note that analogies from conceptu-

ally different domains result in more creative, original ideas. The success of many biologically inspired designs supports that biology is a good source of analogies. However, designers who want to use biological analogies are generally limited by their knowledge of biology. We believe that a systematic search of biological phenomena relevant to a specific design problem will identify a greater variety of potential analogies and likely result in more creative design than simply using analogies that come to mind. We have chosen to take advantage of the enormous amount of biological information already available in natural-language format, such as books, journals, etc. Thus, we developed a method that uses natural language processing to extract relevant biological phenomena from these existing sources of biological knowledge.

Another approach to biomimetic design is to create a database of biological phenomena organized by engineering function (Vincent & Mann, 2002; Lindemann & Gramann, 2004). However, compiling and updating a suitably expansive database is resource intensive and may be subject to the compilers' own knowledge and bias. Lindemann and Gramann (2004) compiled a "checklist" that translates between biology functions and engineering terms based on their own knowledge and current biology sources.

Other challenges of compiling a comprehensive database include the explosive information growth occurring in the biological sciences (Rebholz-Schuhmann et al., 2005), and the dynamic nature of biological knowledge (Spasic et al., 2003). The bioinformatics community recognizes these challenges and is also investigating solutions other than databases, for example, natural language processing and text mining, as alternatives for knowledge discovery and retrieval.

Conceptual design is a creative process, but also a process that should be systematic and intelligent (Dym & Little, 2000) such that appropriate concepts can be generated when required. Our approach to biomimetic design enables the creativity that comes from cross-domain inspiration while providing the systematic and intelligent framework found within the structure of natural language.

Our method involves searching for instances of functional keywords in the biology knowledge source. Matches containing keywords are examined for relevant biological phenomena that can be applied toward the engineering problem of interest. Our initial biological corpus is an introductory university-level textbook (Purves et al., 2001). Other texts can be added or substituted as appropriate for the initial search. The more challenging task is the initial identification of relevant phenomena, whereas locating further details as required on the relevant phenomena is a much more familiar research task.

We obtain our initial keywords by performing a functional decomposition on the engineering problem. Verbs are used to formulate keywords because they convey functionality (Stone & Wood, 2000; Ullman, 2003) and are important to the interpretation of sentences (Joanis & Stevenson, 2003). Using verbs rather than nouns to search the corpus enables the designer to find new biological phenomena that perform related functions. For example, searching for "kidney," which is known to remove toxins from blood, will only provide matches with kidney whereas searching for "remove" will identify other entities that remove (as agents) as well as entities that are removed (as objects). Variation in form that provides the same functionality is discussed by McAdams and Wood (2000) in their work on design by analogy.

Once the relevant biological phenomena are found, designers must apply analogical reasoning to transfer knowledge from the source domain, that is, biology, to the target domain, that is, engineering. Much work has been performed on the application of analogical reasoning to problem solving, most notably Gentner's (1983) structure-mapping theory. Gentner postulates that for between-domain analogies to be use-

ful, there must exist a one-to-one mapping between the domains, and similarities between the domain relations, as opposed to similarities between attributes only. For example, the flow of electrons in a circuit is similar to the flow of people in a subway tunnel, although people do not resemble electrons (Gentner & Holyoak, 1997). Although the analogical reasoning process is crucial to our approach to biomimetic design, this paper only addresses the retrieval of relevant phenomena from the source domain. For details on analogical reasoning for biomimetic design, see Mak and Shu (2004a, 2004b). Past case studies using this method of biomimetic design include those in design for remanufacture (Vakili & Shu, 2001; Hacco & Shu, 2002) and in microassembly (Shu et al., 2003, 2006).

Our natural-language processing approach introduces another challenge: information retrieval between different domains is hampered by differences in domain vocabularies, or, *lexicons* (Hon & Zeiner, 2004; Lindemann & Gramann, 2004). As each domain has its own "sublanguage" (Friedman et al., 2002), a keyword that is used by an engineer may not be useful in biology, resulting in few or no matches in a biological corpus. For a problem that involved cleaning, the keyword "clean" itself resulted in only four relevant matches in the *Life* text. When a biochemist was asked to suggest keywords for the problem of cleaning, he suggested "defend," because cleaning is often performed as a defensive mechanism (Waygood, 2003). To most engineers, defend is not intuitively relevant to cleaning, and defend and clean are not directly related through lexical relationships, for example, as synonyms or antonyms (Manser, 2004; WordNet 2.0, n.d.). Figure 1 below shows a possible lexical path between defend and clean, but it is indirect and relies on two lexical references.

Defend was used in past work (Mak & Shu, 2004a, 2004b) to retrieve many relevant biological analogies for cleaning. The following is one such example from Purves et al. (2001):

When pathogens pass these barriers, plant defenses are activated. Plants seal off and sacrifice the damaged tissue so that the rest of the plant does not become infected. This approach works because most plants can replace damaged parts by growing new stems, leaves, and roots.
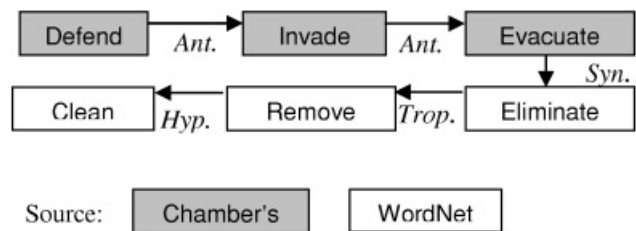

Fig. 1. A possible lexical path between defend and clean.

The above excerpt was presented to engineering students who were asked to use the biological phenomenon to generate concepts to enable clean, dirt-free clothing. The majority of students were able to successfully develop analogy-based concepts, including concepts for modular clothing, where dirty layers or sections are removed and replaced (Mak & Shu, 2004*b*). Although defend is not lexically related to clean or remove, the concepts are biologically related.

This paper describes how we bridge cross-domain terminology when searching biological knowledge in natural-language format to support biomimetic design. We show how to objectively identify useful, but not obvious, keywords that may not be lexically related, for example, as synonyms and antonyms. The method presented also does not rely on expert assistance.

Below we define the nomenclature used in this work and provide a background on language and linguistics, computational linguistics, bioinformatics, and language and design. Following the background section, we illustrate the method used to automatically identify nonobvious keywords using three examples. We start by detailing the clean/remove example already introduced, and then summarize the results for examples in encapsulation and microassembly.

2. NOMENCLATURE

Agent: performer of verb, for example, Pat in "Pat threw the ball." The term agent is used instead of the term subject, to denote the doer, as passive sentences may lack an explicit subject.

Biologically connotative: word not part of a biology term defined in either biology reference below, but appears in definitions of biology terms.

Biologically meaningful: either biologically connotative or biologically significant as defined above and below.

Biologically significant: word identified as part of biology term defined in either Oxford Dictionary of Biology (Hine & Martin, 2004), or biology-online.org (BioOnline, 2004).

Bridge verb/word: verb other than keyword verb that is modified by words frequently collocating with keyword.

Collocation: the occurrence of a word in association with another word, usually the keyword used for searching. Also referred to as a co-occurrence.

Corpus: a written sample of language used for linguistic analysis.

Grammar function: the relationship of a noun phrase to the verb as either an agent/subject or an object (Trask, 1999).

Hypernym: describes the superset of a word, where the hypernym encompasses all instances of *X*. For example, tree is the hypernym of maple (Miller et al., 1993).

Hyponym: describes the subset of a word, where the hyponym is a specific instance of *Y*. For example, tree is a hyponym of plant (Miller et al., 1993).

Keywords: used to search for text documents or passages that contain instances of these words.

Noun phrase: a phrase based on a noun, for example, "nasal *glands*."

Object: receiver of verb action, for example, ball in "Pat threw the ball."

Oblique object: indirect object or object of prepositional phrase, for example, me in "He threw me the ball" and "He threw the ball to me" (Trask, 1999).

Part of speech: also known as lexical category, for example, nouns, verbs, adjectives, and so forth.

Phrase: a linguistic unit consisting of more than one word but does not comprise a complete sentence (Matthews, 1997). See *noun phrase* above for an example.

Prepositional phrase: a phrase that starts with a preposition, for example, "*across* the street."

Sense: the meaning of a word. Words may have multiple senses or meanings. Senses in WordNet are enumerated.

Troponym: specifically refers to the hyponym relationship between verbs. The relationship formula between two verbs is V1 is to V2 in some particular manner (Fellbaum, 1993). For example, "to amble" is a troponym of "to walk" because ambling is a particular manner of walking.

Verb phrase: a phrase based on a verb, for example, "*excrete* salt."

3. BACKGROUND AND LITERATURE REVIEW

Our approach was developed based on work from many diverse fields: linguistics, library and information sciences, natural language processing and computational linguistics, data and text mining, and design theory and methodology. In this section, we will expand on the aspects of each field relevant to the development of our approach to the biomimetic design process.

Although the comprehension of language is a skill fundamental to humans, it is a complex process that is not fully understood. Language, although complex, is governed by a set of rules. Automating the use of information in natural language format requires that computers understand language, and this involves problems studied by the computational linguistics community. Because language is governed by "grammar," or a set of rules, it is possible to algorithmically process language to identify patterns and extract information. In the simplest model of the English language, we recognize that there are verbs and nouns (or more generally, multiple-word noun phrases, e.g., "black cat"). The grammar function of these noun phrases can be either the subject or the object of the verb, and the typical construction of English sentences is subject–verb–object (SVO). We use this language model to identify our "bridge verbs" to connect biology and engineering lexicons.

Early work on quantifying language in the information and library sciences examined word frequencies. This research suggested that word use follows a distribution such that it is possible to determine the frequencies of the most

meaningful words (Zipf, 1949; Luhn, 1959). Word frequencies reflect the author's treatment of the subject matter, as an author will typically use the same words repeatedly to convey a single idea. Zipf's law of language states that there are a few words that occur frequently and many more words that occur infrequently such that it is possible to compute the frequencies of the most important words (Cleveland & Cleveland, 1990). Zipf's law is a specific example of the power law as applied to language (Adamic & Huberman, 2002). We use word frequencies extensively to identify meaningful words in our method.

In latent semantic indexing (LSI), related concepts are revealed across large sets of documents by examining these documents for similar terms. "Latent semantics" refers to how this technique does not require knowledge of the language and its structure, but still extracts document meaning based on word similarity and frequencies across large sets of documents (Deerwester et al., 1990; Landauer & Dumais, 1997; Yu et al., 2003). We identify biological phenomena relevant to engineering problems by looking for meaning-consistent usage of chosen keywords and related alternative keywords.

The limitation of frequency analysis is that it does not account for the *meaning* of words and their relationships to other words. Resolving differences in multiple word meanings, or word sense disambiguation, is a current problem in computational linguistics. For example, the verb "to draw" can mean either "to extract," for example, water, or "to produce a drawing or picture." Examining word collocations, that is, pairs or groups of words occurring together, helps to resolve this ambiguity (Yarowsky, 1995; Banerjee & Pedersen, 2003). Yarowsky's (1995) hypotheses on collocation are the following:

1. Words have one sense per collocation.
2. Words have one sense per discourse.

We rely on these hypotheses to recognize relevant collocations found within our matches, and to assure the coherence of collocations drawn from related matches.

Another related problem being examined by computational linguists, for example, Melamed (2000), concerns the translation of text between different languages. Although we work within the English language, we aim to retrieve similar concepts expressed in a different lexicon, or "sublanguage" of a specific domain. In cross-language information retrieval, the question of whether to translate the query or the document arises. McCarley (1999) found that a hybrid translation approach incorporating translations in both query and document was the most promising. However, in our work, it is more practical to translate, or rather rephrase, the query, by using more biologically meaningful keywords.

Researchers in information-intensive fields are developing data- and text-mining methods to find relevant information quickly and to structure data meaningfully. Although data and text mining do not necessarily rely on language understanding, patterns are identified and models are extracted from data and text that are not in structured databases (Witten & Frank, 2000). This work is relevant to us because we do not want to rely on the availability of structured databases developed specifically to support biomimetic design.

In molecular biology, data mining is used to handle the enormous amount of data produced from projects such as sequencing the human genome (Iliopoulous et al., 2001; Chen et al., 2005; Korbel et al., 2005). A large volume of information is also being produced in biological topics such as diseases, organisms, and protein structures (Rebholz-Schuhmann et al., 2005). Many researchers (Spasic et al., 2003; Chen et al., 2005; Korbel et al., 2005) are working to extract information from MEDLINE, a database of over 14 million articles, spanning 65 years of publications (Rebholz-Schuhmann et al., 2005).

Data mining applications in engineering include monitoring of manufacturing systems such as those in steel mills (Korpipaa, 2001) and textile plants (Ehrenman, 2005). Many engineering applications of data mining typically involve analyzing not text, but numeric data, which is the type of data normally produced from monitoring and quality control processes. Text mining has been used to support decision making in mission-critical aircraft maintenance operations (Farley, 2001).

Engineering design is itself information intensive, as information is required to support decision making at many points of the design process (Gero et al., 1994; Dentsoras, 2005). It is especially important to have access to relevant and meaningful information early in the design process, at the conceptual design stage, as decisions made in this stage have the highest impact on the rest of the process. The added challenge of identifying relevant information from a different domain, that of biology, further motivates us to develop a systematic biomimetic search approach rather than relying on personal knowledge of biology. Our method aims to provide as many as possible relevant biological analogies for the early stages of design, while the possibility still exists for widely different concepts to be developed and considered.

Applying language and language analysis tools to the design process is not a new idea. Many have used language to analyze design results and to model design, incorporating tools such as frequency and collocation analysis (Yang & Cutkosky, 1997), LSI (Dong et al., 2003) and the general concept of "grammar" (or a set of rules) to define permissible combinations (Li & Schmidt, 2000). Language is used to support the design process at different stages and in different fields. Applications of natural language processing exist in mechanical design (Stone & Wood, 2000), software (Burg, 1997), and civil engineering (Yang et al., 1998), as well as in architectural design (Segers, 2004).

Stone and Wood (2000) developed and later worked toward formalizing (Hirtz et al., 2001) a *functional basis*, where the design functionality is described in a VO format.

The verb represents the function, or operation, whereas the object represents the flow. They defined a standard vocabulary to obtain consistent terminology and levels of detail, with the ultimate goal of creating standardized entities for design repositories. Expanding on the functional basis, Sridharan and Campbell (2004) developed a grammar for use with Stone and Wood's design language (2000). The set of grammar rules define permissible interactions between the functions described by the functional basis. Grammar, as a rule set, has been applied to designing gears trains (Li & Schmidt, 2000) and clock mechanisms (Shea & Cagan, 1999; Starling & Shea, 2002).

Yang et al. (1998) generated design thesauri by examining electronic design notebooks to capture and reuse design information. Mabogunje and Leifer (1997) examined distinct noun phrases used in design project documentation and found a correlation between design success and a larger number of distinct noun phrases. Yen et al. (1999) used audio recordings in addition to design documentation and notebooks to capture the knowledge contained in the work in progress that may not be included in final documentation. Dong et al. (2003) used latent semantic analysis to determine the cohesiveness of design team documentation as a measure of success of the design team.

An architectural application of language analysis involves collecting architects' annotations on design sketches to serve as the basis for further design stimuli (Segers, 2004). Segers (2004) and Dentsoras (2005) note the benefits of using words as stimuli at the conceptual design stage, because words are not necessarily tied to a specific physical representation, and thus may be form independent.

Although there is much work in examining language as expressed during design, there has not been as much work on using natural language to generate and facilitate design, with the exception of Segers' work (2004). However, she does not initiate the design process with natural language, but rather uses it as feedback by utilizing the designer's own initial annotations on a sketch. Our method relies on the use of language as a starting point for concept generation through the biomimetic design process.

Although language and linguistics do not traditionally comprise formal engineering and design tools, the variety of language and design applications outlined above establishes that language can be used to support the design process at different stages. In this paper, we show how language analysis can be used to aid conceptual design by retrieving biological phenomena relevant to engineering problems. Specifically, we developed a method for bridging lexicons from different domains to facilitate the retrieval of cross-domain analogies for design.

4. METHOD

We first give a general overview of our bridging method, and then illustrate it using a detailed example that involves cleaning or removing dirt, for example, to enable clean clothes. For this example, we confirm that we can algorithmically generate the bridge verb defend that was first provided by the biochemist, as one of many similarly nonobvious, but biologically meaningful keywords.

In our results and discussion sections, we will summarize two other examples: one that involves encapsulating pigments to improve stability, where biological analogies to encapsulating or enclosing were sought, and another that involves the handling of microobjects.

4.1. Method overview

The corpus was searched for keywords related to the original engineering functions as well as alternative keywords. Alternative keywords are generated using the troponym/ hypernym feature of WordNet. WordNet is an online lexical database that is organized according to current psycholinguistic theories on how people use and remember language, not alphabetically like dictionaries. Troponyms are subordinate verbs describing specific actions, that is, sauntering is a specific way of walking, whereas hypernyms are superordinate verbs.

The resulting matches were examined for relevance. Matches deemed relevant were saved and the words contained in keyword-match passages were counted to identify words that frequently collocated with, or occurred in the vicinity of, search words. High-frequency words were classified as modifying the keyword, modifying another verb, or word with another usage or part of speech. Modifying is used to mean how the frequent word was used relative to its verb, that is, as an agent, object, or oblique object. High-frequency words were often found to be modifying verbs other than the searched keyword. These modified verbs were identified and added to the set of *bridge verbs*.

To objectively determine biological meaningfulness, the bridge verbs were compared to terms defined in two biology dictionaries (BioOnline, 2004; Hine & Martin, 2004). Verbs and their forms that are contained within biological terms are designated as *biologically significant*. Examples of such verbs include "reduce," "protect," and "infect," forms of which appear in the terms "reduction," change in atomic composition through the addition of electrons (BioOnline, 2004); "cryoprotectant," substance that protects tissues from freezing (Hine & Martin, 2004); and "infection," invasion and multiplication of microorganisms in body tissues (BioOnline, 2004).

However, the inclusion of a word in a biology-dictionary term was found to be too limiting of a criterion for biological meaningfulness. Many seemingly meaningful words were used within definitions but were not contained in the terms themselves. One such example is defend, the keyword suggested by the biochemist. Forms of defend, for example, "defense," were used in 27 definitions for terms such as "autoimmunity" and "phagocytes" (Hine & Martin, 2004). The appearance of defend within such definitions indicates a relationship between the defensive functionality

and the immune system, but this relationship is not explicitly expressed in the terms themselves. The words not contained in terms themselves, but do occur in definitions of biological terms, are designated *biologically connotative*.

As word frequencies tend to reflect how authors think about their subject matter, the frequency of bridge verb occurrence in both terms and definitions of terms in a biological dictionary further delineated the most potentially useful words. The bridge verbs were therefore sorted according to descending dictionary count and the occurrences of biologically significant words in this sorted list were summed over the entire set of verbs. Plotting the cumulative density (occurrence) of biologically significant verbs against the logarithm of dictionary count, we observe the dictionary count range that contains the largest number of biologically significant words coincides with the location of the steepest slope of this plot. In this range, we find the words that are not necessarily biologically significant, but also those that are biologically connotative, serve as promising candidate keywords for the next iteration of the search. The graphs produced from this ranking process are shown in the Results section for the examples described next.

4.2. Detailed example and results for remove

The example below is on the problem of cleaning, that is, removing dirt. We first identified troponyms for the keywords clean and remove. Of the 179 troponyms for remove, 38 resulted in one or more matches in the corpus. Figure 2
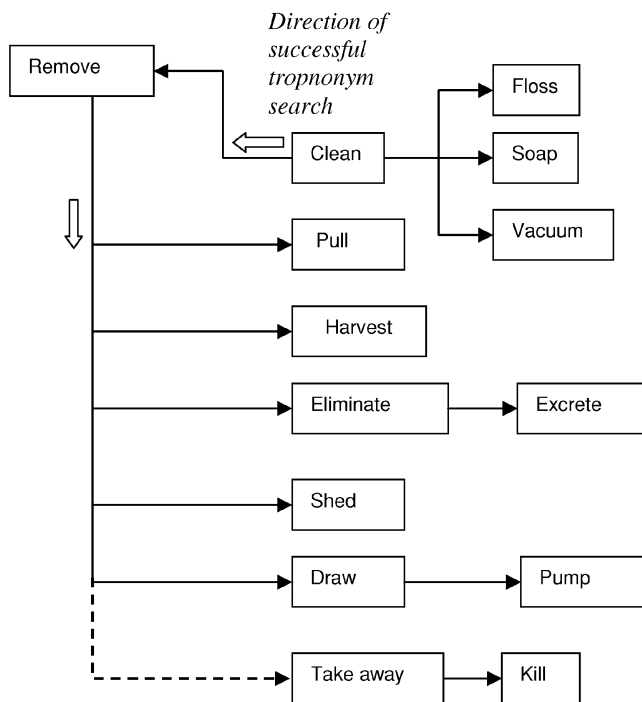

Fig. 2. The WordNet hypernym/troponym search results.

shows the troponyms that ultimately produced more than 10 matches within the biology text. It also illustrates the WordNet hierarchy and how the original keyword clean relates to the successful keywords.

Using troponyms of the keyword enabled us to generate alternative keywords that improved the quantity and quality of matches over those obtained using synonyms. One reason is that WordNet has more troponyms than synonyms. Using troponyms provided some keywords not obviously related to the initial keyword of clean or remove, for example, "excrete," "eliminate," "kill," and "draw" (as in to draw water through capillary action; WordNet 2.0, n.d.). *Life* (Purves et al., 2001) was searched for occurrences of all troponyms of remove, nine of which returned a significant number of matches.

Below are two typical matches for the keywords used. Match 1 is retrieved by the keyword kill. Match 2 is retrieved by the keyword eliminate. The keyword kill resulted in a total of 91 matches in the corpus, and eliminate resulted in a total of 45.

4.2.1. Match 1: Retrieved with the keyword kill

As the virus *kills* more and more TH cells, the immune system is less and less able to defend the body against various diseases (Purves et al., 2001).

4.2.2. Match 2: Retrieved with the keyword eliminate

. . . kangaroo rats reduce populations of some rodent species and *eliminate* others from places where they live. Kangaroo rats compete with other seed-eating rodents both by reducing their food supply—exploitative competition—and by aggressively defending space—interference competition (Purves et al., 2001).

Matches were examined for relevancy with the following cases discarded immediately, as it was found previously that these types of matches tend to be irrelevant (Hacco & Shu, 2002).

1. Nonverb instances of keywords where the nonverb form of a word has a significantly different meaning from the verb form: For example, the verb "strip," another troponym of remove, means to "deprive, divest" and the noun strip means "a relatively long narrow piece of something" (WordNet 2.0, n.d.). Therefore, all noun instances of strip were removed.
2. Verbs acting on "abstract" objects: Because we generally deal with physical phenomena, instances where the verb acts on an abstract entity are removed from the search results. For example, for the troponym eliminate, instances referring to "eliminating risk" were discarded.
3. Keyword being used in a different sense: Verbs are highly polysemous, or have multiple meanings. This is perhaps the most difficult type of irrelevant match

to identify and there is much work addressing this challenge (Yarowsky, 1995; Resnik, 1997). For example, the troponym draw can be used in the sense, "to draw out, or remove," for example, water, or it can mean "to make or trace a figure" as in to draw a diagram (WordNet 2.0, n.d.). Matches with incorrect senses of the keywords were removed.

All remaining matches were saved, and high-frequency words were identified through frequency analysis. Words are deemed "frequent" if their frequency is above an established cutoff. The cutoff was determined by using the 0.01 or 0.025 critical value for a chi-squared distribution with 1 degree of freedom. A cutoff is necessary, as there are many unique words that occur infrequently. The selected cutoff corresponds to approximately the top 1 or 2.5% of the unique word occurrences. Many high-frequency words are nouns, whereas verbs are rarely identified as frequent words, mainly because there are three times as many nouns as verbs contained in the English language, according to the *Collins English Dictionary* (Fellbaum, 1993).

Next, collocation analysis identified the nonkeyword verbs associated with the high-frequency words. For example, the troponym kill produced 91 matches in *Life* (Purves et al., 2001). Analyzing all 91 matches yielded the frequent words "cell," "body," and "disease," among others. Examining Match 1 shown above and again below, it can be seen that "cells" modifies kill as an object, and body modifies defend as an object. "Diseases" modifies defend as the object of a prepositional phrase. Match 1 is shown below with the original keyword kill in bold, frequent words italicized, and the nonkeyword verb defend modified by frequent words in bold and underlined.

> As the virus **kills** more and more TH *cells*, the immune system is less and less able to **defend** the *body* against various *diseases*.

Match 2 is more typical of scientific and technical writing, and is correspondingly more complex. A frequent word produced by the keyword eliminate is "competition," which is italicized in the match given again below.

> . . . kangaroo rats reduce populations of some rodent species and **eliminate** others from places where they live. Kangaroo rats compete with other seed-eating rodents both by **reducing** their food supply—exploitative *competition*—and by aggressively **defending** space—interference *competition* (Purves et al., 2001)

The dashes preceding "exploitative competition" and "interference competition" above are to be read as prepositions, for example, "by reducing their food supply *through* exploitative competition," rendering the frequent word "competition" the oblique object of prepositional phrases "by reducing . . ." and "by aggressively defending. . . ." Thus,

the verbs modified by competition as an oblique object are verbs other than the keyword eliminate.

The two new verbs defend and reduce, identified in the above matches with the troponym keywords kill and eliminate, respectively, are then added to the set of bridge verbs. Bridge verbs collected using this method characterize the retrieved biological phenomena and are related to our original keyword clean/remove.

The bridge verbs generated were then compared against two biology dictionaries (Hine & Martin, 2000; BioOnline, 2004) to determine their *biological significance*. Although the dictionaries typically define noun terms, verbs derivationally related to their noun forms (e.g., protect vs. protection) were designated biologically significant. As described earlier, the *biological connotation* of a verb was measured by its word frequency within definitions in the biology dictionary. The number of occurrences of bridge verbs in BioOnline (2005) was determined, and bridge verbs were sorted by descending number of occurrences. Continuing with our two example bridge verbs, defend and reduce occurred 23 and 400 times, respectively, within the biology dictionary. The semilogarithm density plot was then produced to highlight the frequency range of the most relevant bridge verbs to serve as subsequent search words. This plot is shown later for the clean/remove example.

Figure 3 shows a flowchart illustrating the bridging process described above.

5. RESULTS

The results for three examples, starting with the remove/clean problem introduced above, are presented below.

5.1. Results for remove

Of the 288 bridge words for remove, 122 (42.4%) were biologically significant, that is, contained in terms defined in one of the two biology dictionaries selected (Hine & Martin, 2004; BioOnline, 2004). Figure 4 shows bridge verbs for remove on a minimized spreadsheet sorted by dictionary-occurrence count, with rows corresponding to biologically significant verbs shaded. A region marked with darker shading contains the highest concentration of biologically significant words. The goal of Figure 4 is not to show the details of the data, but rather to illustrate overall data patterns in a type of preliminary data analysis known as data visualization (Witten & Frank, 2000). Such visualization enables quick analysis and highlights interesting data trends for further investigation.

The darkest region for the remove data set centers on words that have 70 to 94 occurrences within the dictionary, with the midpoint at approximately 82 occurrences. There are 31 words within this dense region, and 25/31 (80.6%) of these words are biologically significant.

Following the observation that biological meaningfulness may be a function of dictionary-occurrence count, the
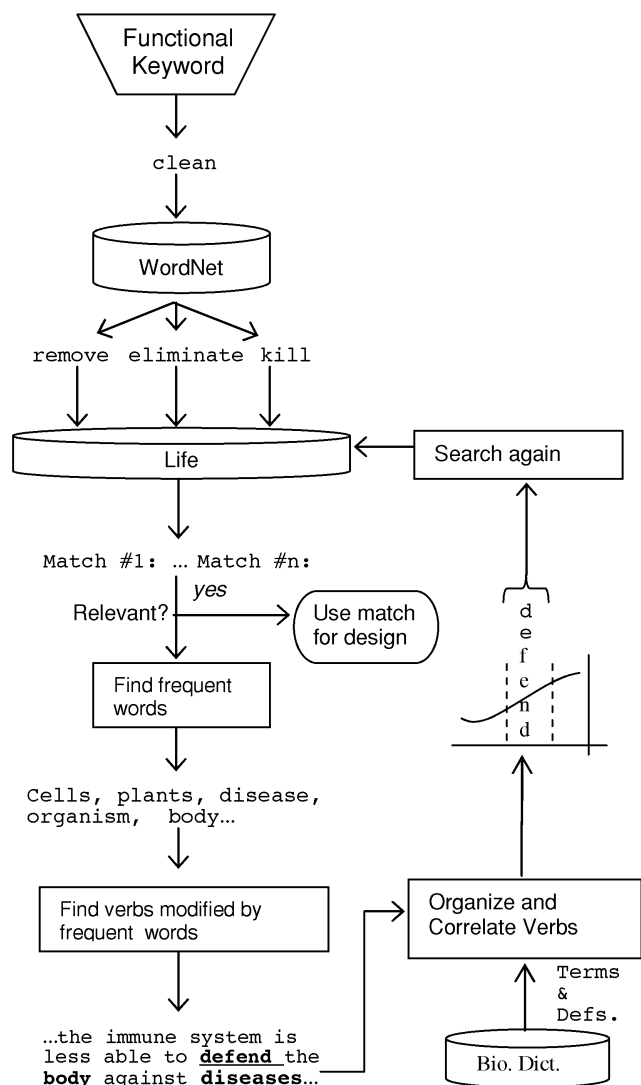
Fig. 3. An overview of the bridging process.

Fig. 4. The densest region of biologically significant bridge verbs for remove.

cally connotative words, for example, defend, within the boundaries. This method is thus able to identify the nonobvious, but highly relevant keyword suggested by the domain expert, along with other similar words, in an objective manner.

Other biologically connotative words for remove are shown in Table 1. Some words in Table 1 appear obviously

Table 1. *Biologically connotative verbs for "remove" data set within boundaries*

Word	Dict. Count
Attach	252
Convert	179
Extract	172
Prevent	163
Surround	151
Spread	129
Inject	77
Pull	68
Maintain	62
Restore	49
Design	48
Conjugate	45
Eject	38
Defend	34
Penetrate	25
Harbor	20
Capture	20

cumulative density of the biologically significant verbs was plotted against the logarithm of the dictionary occurrence counts in Figure 5. In this figure, the steepest part of the curve represents the region with the majority of biologically significant words, with the densest region described above located in the middle of the graph.

In Figure 5, the steepest part of the curve is bounded by 317 and 19 dictionary counts. These boundaries contain 78.7%, or 96 of the 122 biologically significant words of the remove bridge verbs. Of the 185 total words within the upper and lower boundaries, 51.9% are biologically significant.

Although not biologically significant, defend, the search word suggested by a domain expert, occurred 34 times within the dictionary occurrence count, placing it within the range where the majority of biologically significant words are found. The high concentration of biologically significant words suggests the usefulness of the remaining, biologi-
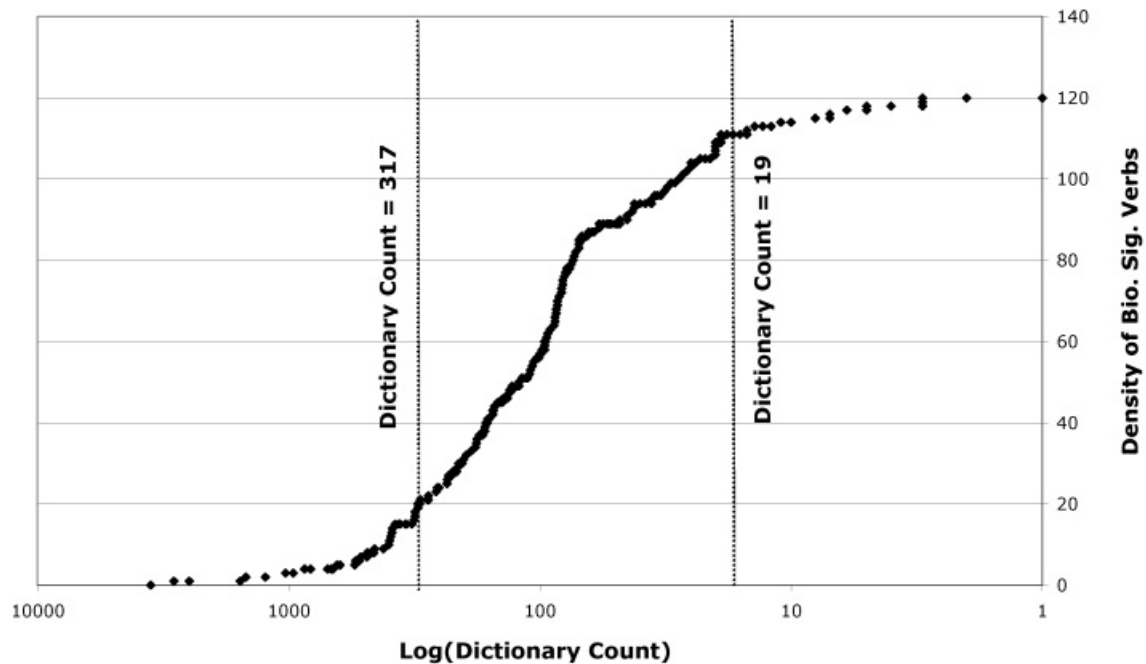
Fig. 5. The density distribution graph for remove.

related to the original keyword. For instance, "attach" can be seen as an antonym for remove, even though this relationship is documented in neither WordNet nor a thesaurus.

Reexamining the matches containing these biologically connotative words along with their collocating keyword(s), a pattern emerges that relates the biological phenomenon, bridge verb, and original keyword. For example, the bridge verb "convert" was retrieved by eliminate as well as other keywords. The matches for convert/eliminate describe the biological phenomena of homeostasis, and the two words can be related using a prepositional phrase:

"Eliminate nitrogenous waste by *converting* it to uric acid."

The above can be rephrased as:

"Convert nitrogenous waste to uric acid to *eliminate* it."

Many useful bridge verbs appear to have this symmetric relationship with the keyword, including defend where

"Organisms *defend* themselves by *removing* parts of themselves,"
or
"organisms *remove* parts of themselves to *defend* themselves."

5.2. Results for "encapsulate"

We have described a method that was able to identify several nonobvious but useful search words, including one that

was suggested by a domain expert for one example. Next, we will confirm that this method can produce useful, biologically connotative words that may not be lexically related to the original search words for other examples. The example described in this section involves encapsulating pigments to improve stability. Thus, biological analogies for encapsulating or enclosing are sought.

"Encapsulate" and "enclose" were the initial keywords. Encapsulate yielded only two matches corresponding to two different biological phenomena. The use of its hypernym enclose yielded an additional 73 matches corresponding to 10 other distinct biological phenomena.

"Rupture," a biologically significant word, is one of the bridge verbs found for encapsulate that occurred in the dictionary 39 times. Although rupture is not listed as an antonym of enclose or encapsulate in thesauri or WordNet, one may be able to draw a pseudo-antonym relationship between rupture and enclose or encapsulate. This method may thus be used to identify relationships not formalized in lexical references.

For encapsulate, there were 76 bridge verbs, 31 (40.8%) of which were biologically significant. Although the encapsulate search was not as exhaustive as the remove search, the quality of the results of the two searches based on biological significance and other lexical properties were comparable. Figure 6 shows the densest region of biologically significant words for encapsulate, which corresponds to 110 to 120 occurrences in the dictionary. In this region, there are four words, all of which are biologically significant.

Figure 7 shows the density distribution plot for the encapsulate data set. It exhibits similar characteristics to the remove

Fig. 6. The densest region of biologically significant bridge verbs for encapsulate.

Table 2. *Biologically connotative verbs for "encapsulate" data set within boundaries*

Word	Dict. Count
Convert	179
Surround	151
Interact	121
Exert	47
Guide	43
Survive	28
Draw	21

data set, with 26 of 31 (83.9%) biologically significant words included within the area bounded by 455 and 18 dictionary counts. Of the total 54 words within the boundaries 26 (48.1%) are biologically significant.

Table 2 lists biologically connotative bridge verbs for the encapsulate problem. This table includes the word "surround," which is, in fact, related to encapsulate as a hypernym. The search for encapsulate was not as extensive as the search for remove, as fewer alternative keywords were generated for encapsulate. The fact that troponyms and/or hypernyms appear automatically within the bridge verb data set suggests that it is not necessary to exhaustively generate alternative search words, or to exhaustively search with them at the outset.

Of the words in Table 2, "survive" is probably the most intuitively biological connotative. Therefore, we will examine some of the matches found in Purves et al. (2001) by searching for forms of survive.
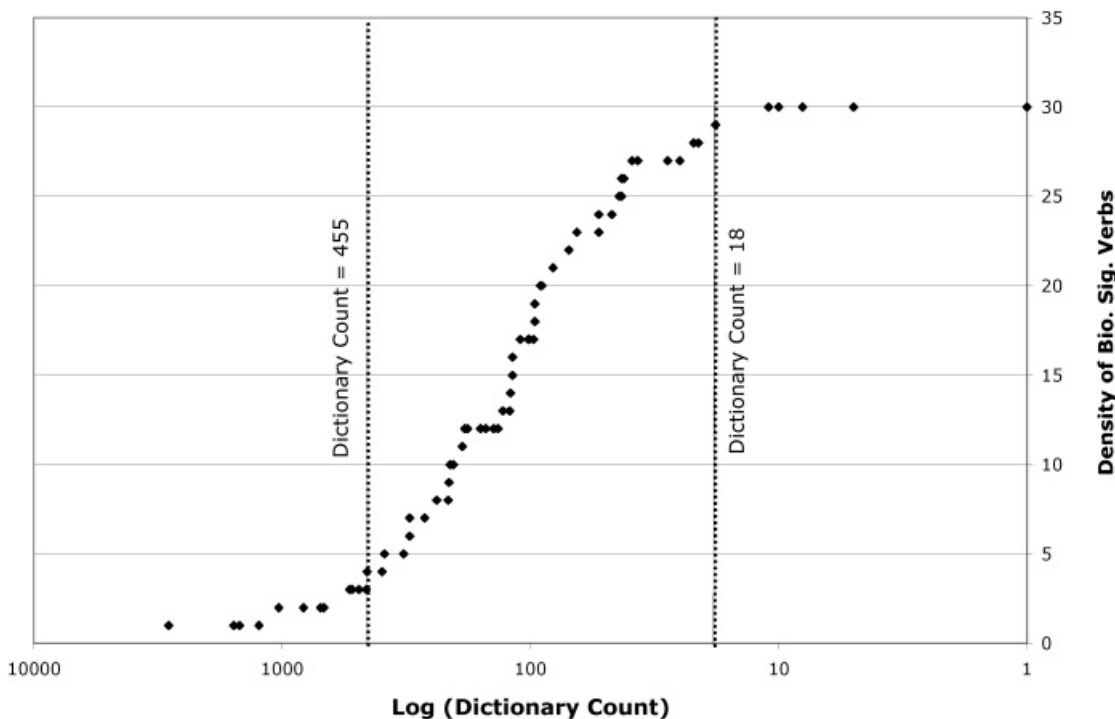

Fig. 7. The density distribution graph for encapsulate.

Many prokaryotes produce no capsule at all, and those that do have capsules can <u>survive</u> even if they lose them, so the capsule is not a structure essential to cell life.

To address this problem, some organisms simply change the lipid compositions of their membranes, replacing saturated with unsaturated fatty acids and using fatty acids with shorter tails. Such changes play a part in the <u>survival</u> of plants and hibernating animals and bacteria during the winter.

The endospore can <u>survive</u> harsh environmental conditions, such as high or low temperatures or drought, because it is dormant—its normal activity is suspended.

The seeds of fireweed not only <u>survive</u> fires, but are encouraged by high temperatures to break their dormancy and sprout.

Eventually, the diploid organism produces thick-walled resting sporangia that can <u>survive</u> unfavorable conditions such as dry weather or freezing.

The above matches all involve some sort of encapsulation, that is, capsule, membrane, spore/seed coating, sporangia, that enhance survival of biological entities. By searching for the biologically connotative word survive, phenomena that involve encapsulation are found. The bridge verb survive and keyword enclose (or encapsulate) also exhibit the symmetric relationship previously mentioned, which is demonstrated below:

"Enclose to survive"
or
"Survive by enclosing"

The relationship between survive and encapsulate/enclose is similar to that for defend and clean/remove. That is, encapsulation/enclosure is performed to enable survival, just as cleaning/removal is performed to enable defense.

5.3. Results for "release"

Our third example involves applying the biomimetic design method to the handling and release of microparts. The handling of microobjects presents different challenges from the handling of macroobjects. At the micro scale, electrostatic, Van der Waals, and surface tension forces dominate gravitational forces, complicating the handling and release of objects.

The details of the physical solution developed to handle and insert a 0.6-mm diameter screw by a conventional robot and end effecter are addressed by Shu et al. (2006). This section presents the use of the bridging and ranking process that produced the keyword and biological phenomena, which led to the physical solution that was implemented.

The initial keyword selected for this problem is release. Although object release is the final step for macroassembly, it is rarely a challenge, as it is for microassembly because

of the domination of the surface forces mentioned above, which cause the microobject to stick to the gripper. Using release to initiate the bridging process, 124 words were retrieved, with 70 (56.4%) of those words being biologically significant. Biologically significant verbs that appear to be relevant include "transfer," "fuse," and "bind." The 124 words from the bridging process were then correlated with their dictionary count from BioOnline (2005), and sorted according to descending dictionary count. Next, the density function of biologically significant words was calculated and plotted against the log of the bridge words' dictionary count, as shown in Figure 8.

Within the boundaries shown in Figure 8, there are 88 words, 56 (63.6%) of which are biologically significant. Table 3 lists the words within the steepest part of the slope that are biologically meaningful, that is, either significant or connotative.

The word "break" is not related to release through any direct lexical relationships (WordNet 2.0, n.d.), and does not appear to be an obvious keyword. Searching Purves et al. (2001) using break as a keyword revealed that it was often retrieved as part of the word "breakdown." In the biology dictionary, breakdown was often found in the two-word noun phrase "break down." Although previous examples of bridge verbs had been single words, noun phrases such as break down suggest the need to consider multiple-word bridge phrases in the future.

Using breakdown to search Purves et al. (2001) located the biological phenomenon of abscission. Abscission is the process by which leaves, petals, and fruits separate from a plant. Abscission is initiated when a growth hormone called auxin is no longer produced, allowing for the further expression of abscisic acid and ethylene, causing parts of the stalk to break down (Purves et al., 2001). Some leaves have a special layer of cells that comprise an abscission zone, which facilitates the breakdown and detachment of the leaf.

In this case, the bridge verb break down and original keyword release can also be related symmetrically:

"*Release* leaf by *breaking down* the leaf stalk."
or
"*Break down* the leaf stalk to *release* leaf."

Table 3. *Biologically meaningful verbs for "release" data set within boundaries*

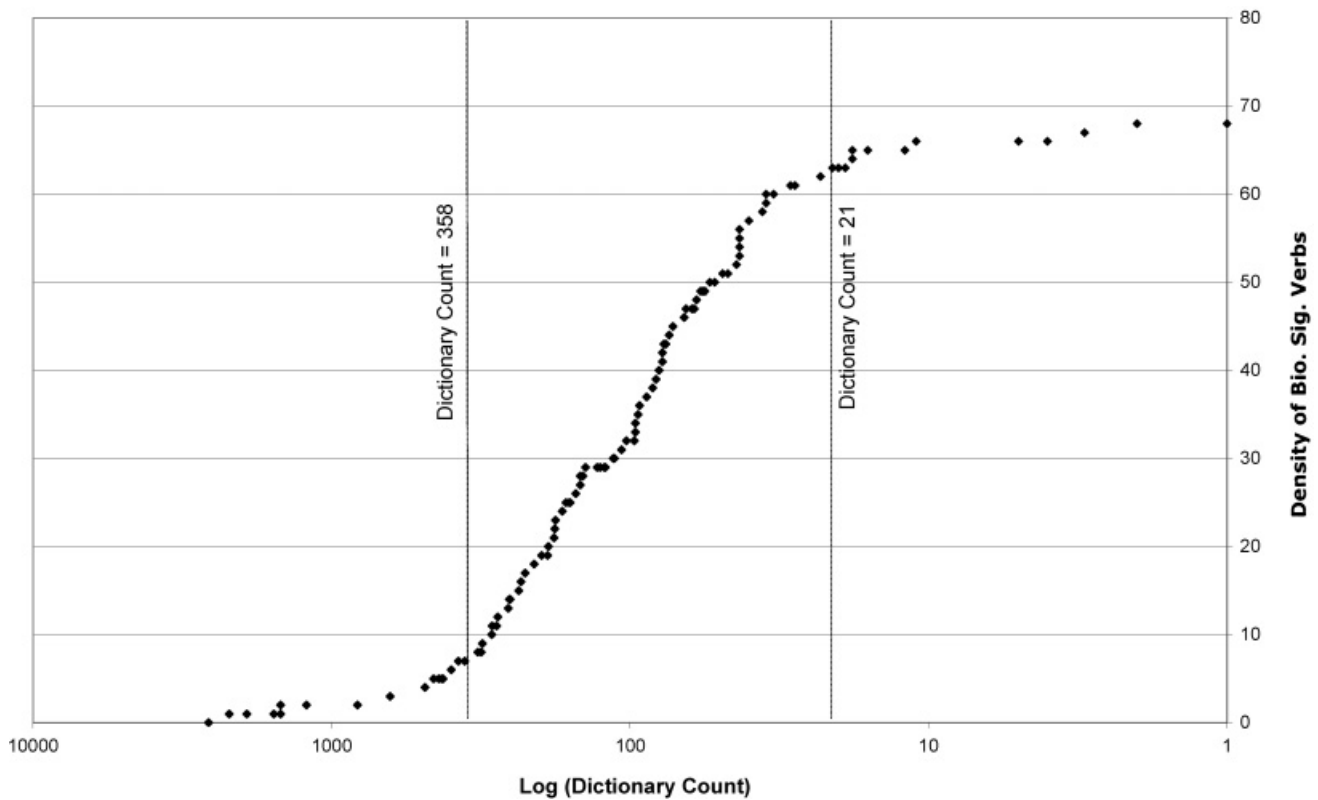| Word	Dict. Count	Biol. Signif.
Continue	152	Yes
Dilate	147	Yes
Flow	147	Yes
Break	144	No
Elevate	141	Yes
Perform	141	No
Spread	129	No

Fig. 8. The density distribution graph for release in the microassembly application.

The concept of an abscission zone could be analogously implemented as an intermediate part between the micropart and gripper, which could, for example, be of significant mass to facilitate its handling by and release from the gripper. This intermediate part may be either left on the micropart or removed subsequently. The actual implementation of this analogy involved heating and pressing the tip of a 4-mm diameter polypropylene rod onto the microscrew, forming a bond between rod and screw. The rod is easily handled and gripped by a typical industrial robot that turns down the screw, which is still attached to the rod. Once the screw is tightened, the resulting increased torque breaks the bond between screw and rod.

Experiments performed at the Technical University of Denmark demonstrated that the abscission analogy retrieved using the bridge word breakdown resulted in a practical concept for handling and releasing microscrews.

6. DISCUSSION

In this section, we begin by discussing the importance of considering both collocation and frequency in the generation of bridge verbs. Next, we expand on the potential of this method to be used as a summarizing method by taking advantage of the simple SVO model of the English language. We then remark on the appropriate uses of lexical references in the bridging process.

6.1. Language analysis to bridge disparate domains

A large number of bridge verbs were generated that include both biologically significant and biologically connotative words. These bridge verbs were sorted by the number of occurrences in the terms and definitions of terms in a biology dictionary. The density of biologically significant words was plotted as a function of dictionary count. These plots showed that over 75% (78.7 and 83.9% for the remove and the encapsulate data sets, respectively) of the biologically significant words can be found within defined boundaries for the two examples. For the release data set, 63.6% of the biologically significant words can be found within the boundaries. For all the data sets, the densest regions of biologically significant words are contained within these boundaries.

The boundaries on the distribution plots enclose a nearly straight-line segment on a semilog plot. The appearance of the line on such a plot is an indicator that Zipf's law of language, and the power law in general, is observed in our data. Our data does not conform entirely to Zipf's law/power law because our selective data sets contain only words that collocate with the keywords.

Based on their position among a high concentration of biologically significant words, the remaining, biologically connotative, words within the boundaries may also serve as promising search words. The usefulness of such search words may not be obvious to domain novices, but they may be

equally, if not more, suitable for retrieving relevant biological phenomena. Defend from the clean/remove problem was first suggested by a biology expert, and was used to identify several biological analogies for the problem. Examining words collocating with the original search words in the clean/remove example commenced the process of algorithmically identifying defend as one of the many biologically meaningful keywords without relying on expert insight, nor on lexical references, which do not list all relationships exhaustively.

A simpler method explored to generate bridge verbs was to merely collect all verbs from all matches. However, this method produced statistically fewer biologically significant verbs and more stative verbs, that is, verbs that do not describe an action, but an unchanging state (Matthews, 1997). This suggests the importance of considering collocation with search words as well as frequency when seeking bridge verbs. Our results along with the results of others (Yang & Cutkosky, 1997; Banerjee & Pedersen, 2003) who combined collocation and frequency analysis suggest that the combination of the two is a better approach for generating bridge verbs.

For the encapsulate example, survive is a biologically connotative word within defined boundaries that has a similar relationship with encapsulate as defend has with clean/remove. One difference is that survive is a stative verb, much like the verbs "to be" and "to have," as opposed to an action verb such as defend. However, the biological relationship, encapsulating enhances surviving, is parallel to removing enhances defending. For the microassembly example, the relationship between break down and release is also similar to that between defend and clean/remove, and both bridge verbs describe actions rather than a state.

Although the majority of biologically significant and connotative words occur within the boundaries, words outside of the boundaries may also be useful. Words below the lower boundary (i.e., below dictionary count of approximately 20) appear more useful than words that lie above the upper boundary (i.e., above dictionary count of approximately 400), as they are less common. Words above the upper boundary may occur too frequently within the lexicon to return many meaningful biological phenomena. Interesting words below the lower boundary for the remove data set include "deplete," "reject," and "invade," whereas words above the upper boundary include "use," "act," and "call." The words above the upper boundary tend to be frequently used verbs within English (Leech et al., 2001) as found in the British National Corpus, a standard corpus used by computational linguists. Interesting verbs below the lower boundary for the encapsulate data set include "bulge" and "engulf." The relationship between bulge and encapsulate is less obvious, whereas engulf is actually a troponym of enclose, one of the original search words used. Similar verbs for the release data set include "pinch" and "resorb," which could be seen as antonym-like to release.

6.2. Language analysis as a summarizing mechanism

Frequent nouns were often found to be agents or objects of the searched keyword. Further examination revealed that the agents, objects, and keywords by themselves might be used to succinctly describe biological concepts associated with each keyword. Therefore, the most frequently collocated words may also be used to capture the dominant biological theme associated with each of the search keywords, and be used as a summarizing mechanism when many matches are returned. For example, the troponym eliminate is associated with "species" (at 20 occurrences), and refer to how interactions between prey and predator species lead to one another's elimination. The troponym "harvest" is associated with "energy" (at 18 occurrences), whereas the troponym "excrete" is closely associated with "water," with 57 occurrences.

Because of the common use of the passive voice in scientific writing, these frequent words are more likely to be the object than agent of the verb used as the search keyword. We can thus generalize the dominant biological theme using the VO format, where the verb is the search keyword and the object is identified from the frequently occurring words that collocated with the keyword. Therefore, we can answer the question "what is being 'eliminated/harvested/excreted'" and then proceed to find the underlying biological phenomena that "eliminates-species," "harvests-energy," and "excretes-water" to use as a basis for concept generation.

6.3. Use of lexical references

The following implications regarding the use of lexical references such as WordNet were discovered while developing the bridging method.

It is not necessary to exhaustively generate troponyms initially to use as search words, as many troponyms are generated when the bridging process is performed. It is also not necessary to perform an exhaustive search of the corpus using all possible troponyms. Remove had 179 troponyms, of which only 38 produced matches, with only 9 producing 10 or more matches. Comparing the initial set of bridge verbs to the list of troponyms will enable a more targeted search. For the encapsulate example, engulf and surround are lexically related to enclose/encapsulate, and appeared in the set of bridge verbs for enclose/encapsulate. The set of bridge verbs for release includes "secrete." which is an indirect troponym of one of the physical senses of release.

This method overcomes limitations of lexical references and may identify new relationships between words that are not yet formally documented. Although some words generated using this method may seem related to the original search words, often no such relationship has been captured within a lexical reference. "Reduce," a verb from the remove set of bridge verbs appears to be related to remove in a synonymous relationship, in that reduce is like remove but

to a lesser degree. Similarly, "rupture" from the encapsulate set of bridge verbs is the opposite of encapsulating or enclosing, as rupture describes an abrupt separation (WordNet 2.0, n.d.). Another undocumented antonymous relationship is found between the keyword release and one of its bridge verbs "adhere."

Documented lexical relationships depend on the reference chosen. It is possible to consult several lexical references, but this method enables the corpus itself to serve as a guide to the authors' representation of lexical relationships. Although it is helpful to use biological and lexical references, this work suggests how to use them such that they enhance, and not limit our search for information contained in natural-language format.

7. SUMMARY AND CONCLUDING REMARKS

This paper describes a method that enables the design engineer to systematically retrieve biological phenomena relevant to engineering design problems. We developed a natural-language approach to biomimetic design to avoid the immense task of categorizing biological phenomena for engineering purposes. We also chose to take more direct advantage of the enormous amount of existing knowledge in journals, books, and so forth. We encountered differences in engineering and biology lexicons, which present challenges for retrieving information from the biology domain, as engineers may not know the most useful keywords. The problems of differing lexicons are addressed through exploring the corpus itself, thus providing not only an accurate picture of the domain, but of the authors' specific perception of the domain. Using our bridging method, we were able to algorithmically generate a nonobvious keyword provided by a domain expert as one of many other nonobvious but relevant keywords. One characteristic for the usefulness of the bridge word appears to be whether it can be expressed in a symmetric manner with the original keyword in the biological phenomenon that related the two.

We believe that this work is an important contribution to engineering design as it systematizes the retrieval of relevant phenomena outside of the engineer's own domain, thus promoting creative and innovative solutions to engineering problems. Although biology is an especially promising source of relevant analogies, the method described in this paper is not domain-specific. Given appropriate domain references, any domain of interest to the engineer can be bridged to retrieve relevant phenomena.

ACKNOWLEDGMENTS

REFERENCES

Adamic, L.A., & Huberman, B.A. (2002). Zipf's law and the Internet. *Glottometrics 3*, 143–150.

Banerjee, S., & Pedersen, T. (2003). *The design, implementation and use of the Ngram Statistics Package*. Accessed at http://www.d.umn.edu/~tpederse/nsp.html

Benami, O., & Jin, Y. (2002). Creative stimulation in conceptual design. *Proc. ASME DETC/CIE*, Paper No. DETC2002/DTM-34023, Montreal.

Biology Online. (2004). Accessed at www.biology-online.org

Biology Online. (2005). Accessed at www.biology-online.org

Burg, J.F.M. (1997). *Linguistic instruments in requirements engineering*. PhD Thesis. Amsterdam: Vrije Universiteit.

Chen, L., Liu, H., & Friedman, C. (2005). Gene name ambiguity of eukaryotic nomenclatures. *Bioinformatics 21(2)*, 248–256.

Cleveland, B., & Cleveland, A. (1990). *Introduction to Indexing and Abstracting*, 2nd ed. Englewood, CO: Libraries Unlimited.

Deerwester, S., Dumais, S.T., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the Society for Information Science 41(6)*, 391–407.

Dentsoras, A.J. (2005). Information generation during design: information importance and design effort. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing 19(1)*, 19–32.

Dong, A., Hill, A.W., & Agogino, A.M. (2003). A document analysis method for characterizing design team performance. *Journal of Mechanical Design 126(3)*, 378–385.

Dym, C., & Little, P. (2000). *Engineering Design, A Project-Based Introduction*. New York: Wiley.

Ehrenman, G. (2005). Mining what others miss. *Mechanical Engineering 127(2)*, 26–31.

Farley, B. (2001). Extracting information from free-text aircraft repair notes. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing 15(4)*, 295–305.

Fellbaum, C. (1993). *English verbs as a semantic net. Five Papers on WordNet*, pp. 40–61. Accessed at ftp://ftp.cogsci.princeton.edu/pub/wordnet/5papers.ps

Friedman, C., Kra, P., & Rzhetsky, A. (2002). Two biomedical sublanguages: a description based on the theories of Zellig Harris. *Journal of Biomedical Informatics 35*, 222–235.

Gero, J.S., Sushil, J.L., & Kundu, S. (1994). Evolutionary learning of novel grammars for design improvement. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing 8(2)*, 83–94.

Gentner, D. (1983). Structure-mapping: a theoretical framework for analogy. *Cognitive Science 7(2)*, 155–170.

Gentner, D., & Holyoak, K.J. (1997). Reasoning and learning by analogy. *American Psychologist 52(1)*, 32–34.

Gordon, W.J.J. (1961). *Synectics, The Development of Creative Capacity*. New York: Harper & Row.

Hacco, E., & Shu, L.H. (2002). Biomimetic concept generation applied to design for remanufacture. *Proc. ASME Design Engineering Technological Conf.*, Paper No. DETC2002/DFM-34177, Montreal, September 29–October 2.

Hine, R.S., & Martin, E., Eds. (2004). *A Dictionary of Biology*. New York: Oxford University Press.

Hirtz, J.M., Stone, R.B., McAdams, D.A., Szykman, S., & Wood, K.L. (2001). Evolving a functional basis for engineering design. *Proc. ASME DETC/CIE*, Paper No. DETC2001/DTM-21688, Pittsburgh, PA.

Hon, K.K.B., & Zeiner, J. (2004). Knowledge brokering for assisting the generation of automotive product design. *Annals of the CIRP 53(1)*, 159–162.

Iliopoulous, I., Enright, A.J., & Ouzounis, C.A. (2001). TextQuest: document clustering of Medline abstract for concept discovery in molecular biology. *Pacific Symp. Biocomputing 2001*, pp. 384–395.

Joanis, E., & Stevenson, S. (2003). A general feature space for automatic verb classification. *Proc. 10th Conf. European ACL*, pp. 163–170.

Korpipaa, P. (2001). Visualizing constraint-based temporal association rules. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing 15(5)*, 401–410.

Korbel, J.O., Doerks, T., Jensen, L.J., Perez-Iratxeta, C., & Kaczanowski, S. (2005). Systematic association of genes to phenotypes by genome and literature mining. *PLoS Biology 3(5)*, e134.

Landauer, T.K., & Dumais, S.T. (1997). A solution to Plato's problem: the

latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychology Review 104(2)*, 211–240.

Leech, G., Rayson, P., & Wilson, A. (2001). *Companion Website for word frequencies in written and spoken English: based on British National Corpus*. Accessed at www.comp.lancs.ac.uk/ucrel/bncfreq/

Li, X., & Schmidt, L. (2000). Grammar-based designer assistance for epicyclic gear trains. *Proc. ASME DETC/CIE*, Paper No. DETC2000/DTM-14574, Baltimore, MD.

Lindemann, U., & Gramann, J. (2004). Engineering design using biological principles. *Proc. Int. Design Conf., Design 2004*, Vol. 5, pp. 18–21, Dubrovnik.

Luhn, H.P. (1959). Auto-encoding of documents for information retrieval systems. In *Modern Trends in Documentation*, pp. 45–58. London: Pergamon Press.

Mabogunje, A., & Leifer, L. (1997). Noun phrases as surrogates for measuring early phases of the mechanical design process. *Proc. ASME DETC/CIE*, Sacramento, CA.

Mak, T.W., & Shu, L.H. (2004a). Abstraction of biological analogies for design. *Annals of the CIRP 53(1)*, 117–120.

Mak, T.W., & Shu, L.H. (2004b). Use of biological phenomena in design by analogy. *Proc. ASME DETC/CIE*, Paper No. DETC2004/DETC-57303, Salt Lake City, UT.

Manser, M., Ed. (2004). *The Chambers Thesaurus*. Edinburgh: Chambers Harrap Publishers Ltd.

Matthews, P. (1997). *The Concise Oxford Dictionary of Linguistics*. New York: Oxford University Press.

McAdams, D., & Wood, K. (2000). Quantitative measures for design by analogy. *Proc. ASME DETC/CIE*, Paper No. DETC2000/DTM-14562, Baltimore, MD.

McCarley, J.S. (1999). Should we translate the documents or the queries in cross-language information retrieval? *Proc. 37th Annual Meeting*, pp. 208–214. College Park, MD: Association for Computational Linguistics.

Melamed, D.I. (2000). Models of translational equivalence among words. *Computational Linguistics 26(1)*, 221–249.

Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. (1993). *Introduction to WordNet: an on-line lexical database. Five Papers on WordNet*, pp. 1–25. Accessed at ftp://ftp.cogsci.princeton.edu/pub/wordnet/5papers.ps

Purves, W.K., Sadava, D., Orians, G.H., & Heller, H.C. (2001). *Life, The Science of Biology*, 6/e. Sunderland, MA: Sinauer Associates.

Rebholz-Schumann, D., Kirsch, H., & Couto, F. (2005). Facts from text—is text mining ready to deliver? *PLoS Biology 3(2)*, e65.

Resnik, P. (1997). Selectional preference and sense disambiguation. *Proc. ACL SIGLEX Workshop on Tagging Text with Lexical Semantics: Why, What and How?*, pp. 79–86, Somerset, NJ.

Segers, N. (2004). *Computational representations of words and representations of words and associations in architectural design, development of a system supporting creative design*. PhD Thesis. Technische Universiteit, Eindhoven.

Shea, K., & Cagan, J. (1999). Language and semantics of grammatical discrete structures. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing 13(4)*, 241–251.

Shu, L., Lenau, T., Hansen, H., & Alting, L. (2003). Biomimetics applied to centering in microassembly. *Annals of the CIRP 52(1)*, 101–104.

Shu, L.H., Hansen, N.H., Gegeckaite, A., Moon, J., & Chan, C. (2006). Case study in biomimetic design: Handling and assembly of microparts. *Proc. ASME IDETC/CIE*, Paper No. DETC2006-99398, Philadelphia, PA.

Spasic, I., Nenadic, G., & Ananiadou, S. (2003). Using domain-specific verbs for term classification. *Proc. ACL Workshop on Natural Language Processing in Biomedicine*, Sapporo, Japan, pp. 17–24.

Sridharan, P., & Campbell, M.I. (2004). A grammar for function structures. *Proc. ASME DETC/CIE*, Paper No. DETC2004-57130, Salt Lake City, UT.

Starling, A., & Shea, K. (2002). A clock grammar: the use of a parallel grammar in performance-based mechanical synthesis. *Proc. ASME DETC/CIE*, Paper No. DETC2002/DTM-34026, Montreal.

Stone, R.B., & Wood, K.L. (2000). Development of a functional basis for design. *Journal of Mechanical Design, Transactions of the ASME 122*, 359–369.

Trask, R.L. (1999). *Key Concepts in Language and Linguistics*. London: Routledge.

Ullman, D. (2003). *The Mechanical Design Process*, 3rd ed. New York: McGraw–Hill.

Vakili, V., & Shu, L.H. (2001). Towards biomimetic concept generation. *Proc. ASME DETC/CIE*, Paper No. DETC2001/DTM-21715, Pittsburgh, PA.

Vincent, J., & Mann, D. (2002). Systematic technology transfer from biology to engineering. *Philosophical Transactions of The Royal Society: Physical Sciences 360*, 159–173.

Waygood, E.B. (2003). Personal communication, Coordinator of Health Research, University of Saskatchewan.

Witten, I.H., & Frank, E. (2000). *Data Mining, Practical Machine Learning Tools and Techniques with Java Implementations*. San Francisco, CA: Morgan Kaufmann.

WordNet 2.0 (n.d.). Accessed at http://www.cogsci.princeton.edu/~wn/

Yang, M.C., & Cutkosky, M.R. (1997). Automated indexing of design concepts for information management. *Proc. Int. Conf. Engineering Design*, Tampere, Finland, August 19–21.

Yang, M.C., Wood, W.H., & Cutkosky, M.R. (1998). Data mining for thesaurus generation in informal design information retrieval. *Proc. 1998 Int. Congr. Civil Engineering*, Boston, October 18–21.

Yarowsky, D. (1995). Unsupervised word-sense disambiguation rivalling supervised methods. *Proc. 33rd Annual Meeting of the Association for Computational Linguistics*, pp. 189–196.

Yen, S., Fruchter, R., & Leifer, L. (1999). Facilitating tacit knowledge capture and reuse in conceptual design activities. *Proc. ASME DETC/CIE*, Paper No. DETC99/DTM-8781, Las Vegas, NV.

Yu, C., Cuadrudo, J., Ceglowski, M., & Payne, J. (2003). *Patterns in Unstructured Data, Discovery, Aggregation and Visualization*. Paper presented to the Andrew W. Mellon Foundation. Accessed at http://javelina.cet.middlebury.edu/lsa/out/lsa_explanation

Zipf, G.K. (1949). *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Cambridge, MA: Addison–Wesley.

Ivey Chiu received her BASc in engineering science (manufacturing systems engineering) and her MASc in mechanical and industrial engineering, both at the University of Toronto. She is currently completing her PhD and teaches design courses at the University of Toronto.

Lily H. Shu received her SM and PhD in mechanical engineering at the Massachusetts Institute of Technology. She is currently in the Department of Mechanical and Industrial Engineering at the University of Toronto. Dr. Shu was awarded the F.W. Taylor Medal by the CIRP (International Academy of Production Engineering) for her work on applying biomimetic design to microassembly.