Proceedings of IDETC'05
ASME 2005 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference
September 24-28, 2005, Long Beach, California, USA

DETC2005-84908

BRIDGING CROSS-DOMAIN TERMINOLOGY FOR BIOMIMETIC DESIGN

I. Chiu and L.H. Shu*

Department of Mechanical and Industrial Engineering
University of Toronto
5 King's College Road, Toronto, Ontario M5S 3G8 Canada
*shu@mie.utoronto.ca

ABSTRACT

This work aims to improve creativity and innovation in design by facilitating the use of cross-domain analogies, particularly from biological phenomena, as stimulus for concept generation. Rather than create an enormous database of biological knowledge to specifically support engineering design, we have chosen to take advantage of the large amount of biological knowledge already in natural-language format, e.g., books, journals, etc. Relevant biological analogies for any given design problem are found by searching for instances of functional keywords that describe the intended effect of the design solution in a natural-language corpus.

However, the optimal choice of keywords, or search terms, is complicated by the fact that engineers and biologists may use differing domain-specific lexicons to describe related concepts. Therefore, an engineer without sufficient background in biology may not be able to identify keywords with biological connotation that are not obviously related to the engineering keywords.

This paper describes efforts to bridge the gap in lexicons by examining words that frequently collocate with searched words. The biological meaningfulness of these bridge words is characterized by how frequently they occur within definitions of biological terms in a biology dictionary. Search words identified this way may not be obvious to domain novices, and may parallel those suggested by domain experts, thus facilitating the use of cross-domain ideas to support design.

Our approach of generating bridge words with biological meaningfulness is generic and can be used to bridge any disparate domains (e.g., engineering and economics). Thus designers are enabled to quickly access relevant concepts from different domains to produce more innovative solutions.

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 cocklebурs attach to clothing and fur. Other work includes correlation of heat transfer principles to shapes found in nature to aid in optimization (Bejan, 2000). Benami and Jin (2002) note that analogies from conceptually different domains result in more creative, original ideas. In the development of synectics, Gordon (1961) observed that biology provided the richest source of direct analogies. The success of many biologically inspired designs also supports that biology is a good source of analogies. However, designers are generally limited by their personal knowledge of biology.

One approach to overcome this limitation 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. This and other challenges of database incompleteness are recognized by the bioinformatics community struggling to keep up with an explosive growth of information (Rebholz-Schuhmann et al., 2005).

Our approach is to take advantage of the enormous amount of biological information already available in natural-language format, such as books, journals, etc. Instances of functional keywords are sought in the biological corpus, or body of text. Matches, or text excerpts containing keywords, are examined for relevant biological phenomena that can be applied towards the engineering problem. Our initial biological corpus is an introductory university-level textbook: *Life, the Science of Biology* (Purves et al., 2001). Verbs are used to formulate keywords because they convey functionality (Stone & Wood, 1999; Ullman, 2003) and are important to the interpretation of sentences (Joanis & Stevenson, 2003).

Past case studies using this method include those in design for remanufacture (Vakili & Shu, 2001; Hacco & Shu, 2002) and centering in microassembly (Shu et al., 2003). Fundamental work performed to improve this method includes observations of how descriptions of biological phenomena are used to solve problems (Mak & Shu 2004a, 2004b) and exploration of linguistic approaches for generating alternative keywords (Chiu & Shu, 2004).

One challenge of this approach involves differing terminology (Hon & Zeiner, 2004; Lindemann & Gramann, 2004) as biologists and engineers use domain-specific lexicon to describe their work. When a biochemist was asked to suggest words relating to the engineering 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, e.g., as synonyms or antonyms. However, "defend" was used in past work (Mak & Shu 2004a, 2004b) to retrieve many relevant biological analogies for cleaning. A primary objective of this work is the automatic identification of such relevant, but non-obvious keywords that cannot be identified through lexical relationships, e.g., synonyms.

Automating the use of natural-language biological knowledge requires that computers understand language, and involves problems studied by the computational linguistics community. Because language is governed by grammar, a set of rules, it is possible to algorithmically process it to extract information. Several engineering applications have made use of "grammars" in the extended sense to generate and describe design (Li & Schmidt, 2000; Starling & Shea, 2002; Sridharan & Campbell, 2004). Yang & Cutkosky (1999) and Wood et al., (1998) generated design thesauri by examining electronic design notebooks to capture and reuse design information.

Central to the problem of *relevance* is that words can have multiple meanings, contributing to ambiguity. For example, the verb "to draw" can either mean "to extract," e.g., "to draw water" or to produce a drawing e.g., "to draw a picture." Some work performed to address word sense disambiguity uses collocation analysis, or examining pairs or groups of words that occur together (Banerjee & Pedersen, 2003; Yarowsky, 1995). Latent semantic indexing identifies relevant documents by comparing similar terms (Deerwester et al., 1990).

This paper describes how we overcome the challenges of bridging cross-domain terminology when using biological knowledge in natural language to support biomimetic design. Following definition of terms in the nomenclature section below, we refer to an ongoing example detailed in past work to provide context, as well as to illustrate the usefulness of a non-obvious keyword given by a biology expert. Next, we describe a method for automatically identifying such non-obvious keywords in general, and confirm that the expert keyword is found using this method. Finally, we use the method to identify similar non-obvious keywords for another example.

NOMENCLATURE

<u>Agent</u> – Performer of verb, e.g., Pat in "Pat threw the ball."

<u>Biologically connotative verb</u> – a verb whose forms appear in the definitions of biological terms in a biology dictionary, but not within biological terms themselves.

<u>Biologically significant verb</u> – a verb whose forms appear within a biological term defined in a biology dictionary.

<u>Bridge verbs</u> – verbs collected from keyword-match passages that can be used to bridge engineering and biology domains.

<u>Collocation</u> – the occurrence of a word in association with another word, usually the keyword used for searching.

<u>Corpus</u> – a written sample of language for linguistic analysis.

<u>Hypernym</u> – 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).

<u>Hyponym</u> – 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).

<u>Keyword-match passage</u> – text excerpt containing the keyword.

<u>Lexical relationship</u> – relationship between words, e.g., synonym, antonym, troponym and hypernym, documented in lexical references such as WordNet, dictionaries and thesauri.

<u>Lexicon</u> **–** Vocabulary specific to a domain.

<u>Modify</u> – Describes how the noun relates to the verb, e.g., in "Pat threw the ball," the ball modifies "threw" as an object.

<u>Object</u> – Receiver of verb, e.g., the ball in "Pat threw the ball."

<u>Oblique object</u> – Indirect object or object of prepositional phrase, e.g., me in "Pat threw me the ball" and "Pat threw the ball to me."

<u>Sense</u> – Meaning of a word. Words may have multiple senses, e.g., "draw" in drawing a figure vs. drawing water.

<u>Troponym</u> – Specifically refers to the hyponym relationship between verbs. The troponym relationship 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.

<u>WordNet</u> – Lexical database organized according to current theories of human language memory rather than alphabetically.

PAST WORK AND MOTIVATION

Past work (Chiu & Shu, 2004) focused on generating additional search words to obtain relevant biological phenomena from natural-language text. Since we chose not to rely on the existence of a biological database organized by standardized keywords, the use of additional search words is crucial for increasing the number of biological phenomena identified. For example, when searching for biological analogies for the process of cleaning, i.e., removing dirt, we found that the search word "clean" did not yield many useful matches, but "remove," a hypernym of "clean" did. In addition, a process that acts to remove a substance can also be described as eliminating a substance. Therefore, forms of both "remove" and "eliminate" should be used as search words to locate processes potentially relevant to cleaning.

Past work used WordNet as a language framework, and found that troponyms were superior to synonyms for increasing the quality and quantity of matches. One reason for this finding is that WordNet contains more troponyms than synonyms. In the above, "eliminate" is a troponym of "remove." Troponyms expand the search space, but use of troponyms and any other lexical relationship restricts the designer to only words documented within a lexical reference such as WordNet or a thesaurus and does not give the designer insight into the domain-specific terminology.

When asked for suitable keywords for the cleaning problem, a biochemist suggested "defend" since many biological entities clean themselves as a defensive mechanism. "Defend" does not have an obvious lexical relationship with either "clean" or "remove." Figure 1 shows a possible path between "defend" and "clean/remove," traversing through several intermediary words (WordNet 2.0; Manser, 2004).

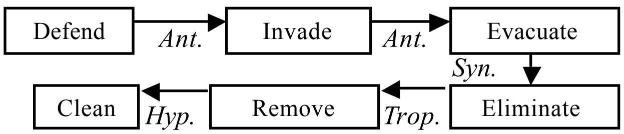

Figure 1: Possible lexical path between "defend" and "clean/remove."

Although not lexically related directly, "defend" and its forms were used to locate several relevant phenomena for cleaning/removing. Two such examples from *Life* (Purves et al., 2001) follow:

> 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.

The above excerpt was presented to engineering students who were asked to apply 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, 2004b). Another excerpt from *Life* (Purves et al., 2001) located by searching for forms of "defend" follows:

> Barriers and local agents **defend** the body – skin is a primary innate **defense** against invasion. The bacteria and fungi that normally live and reproduce in great numbers on our body surfaces without causing disease are referred to as normal flora. These natural occupants of our bodies compete with pathogens for space and nutrients, so normal flora are a form of innate **defense**.

Using the above excerpt as stimulus, many students developed concepts using substances, e.g., coatings to displace dirt (Mak & Shu, 2004b) and prevent it from settling. Although "defend" is not lexically related to clean or remove, the concepts are biologically related. In addition, searching for "defend" located several biological phenomena that served as useful analogies for the clean/remove dirt problem. Therefore, the goal of this work is to automatically and objectively identify such useful, non-obvious keywords.

METHODOLOGY

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 1) modifying the keyword; 2) modifying another verb; or 3) word with another usage or part of speech. "Modifying" is used to mean how the frequent word was used relative to its verb, i.e., as an agent, object or oblique object. High-frequency words were often found modifying verbs other than the searched keyword.

For example, the word "kill" was identified as a troponym of "remove." Counting all the words within the matches retrieved by "kill" found "cells," "body" and "diseases" to be high-frequency words. Purves et al. (2001) illustrate the relationship between these words:

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

In the above, "cells" modifies the keyword "kill" as an object; "body" modifies "defend" as an object and "diseases" modifies "defend" as the object of a prepositional phrase.

Verbs such as "defend" that are modified by frequently collocated words appear to be biologically meaningful as well as related to the original keywords. These verbs will be called *bridge verbs* as they can potentially bridge the gap between engineering and biology terminologies.

To objectively identify their biological significance, the bridge verbs were compared to terms defined in two biology dictionaries (Hine & Martin, 2004; www.biology-online.org, 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 (www.biology-online.org, 2004); "cryoprotectant" – substance that protects tissues from freezing (Hine & Martin, 2004); and "infection" – invasion and multiplication of microorganisms in body tissues (www.biology-online.org, 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," e.g., "defense," were used in 27 definitions for terms such as "autoimmunity" and "phagocytes" (Hine & Martin, 2004). Thus, a relationship exists between the defensive functionality and the immune system, but the 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*.

Library and information science research suggests 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. Therefore, how frequently bridge verbs occurred in both terms and definitions of terms in a biological dictionary further delineated the potentially most useful bridge verbs.

Details for steps of this approach for "remove" follow.

<u>Generate bridge verbs</u> – Search words used for the cleaning, or removing dirt problem are "remove," a hypernym of "clean," and "eliminate," "harvest," "pull," "pump," "shed," "excrete,"

"kill" and "draw," all troponyms of "remove." Frequently collocated words for these keywords were found. Next, the verbs being modified by the frequent words were determined using a part-of-speech tagger (Brill, 1994), a partial parser (Abney, 2002) and a Perl script.

We will continue briefly with "eliminate" as a specific example. Searching the text for "eliminate" produced 45 relevant matches in *Life* (Purves et al., 2001) including:

> ... 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.

The dashes preceding "exploitative competition" and "interference competition" above are to be read as a preposition, e.g., "by reducing their food supply *through* exploitative competition," rendering the frequent word "competition" as the oblique object of prepositional phrases "by reducing…" and "by aggressively defending …". Thus the verbs modified by "competition" as an oblique object are "defend" and "reduce." "Defend" and "reduce" are then added to the set of bridge verbs.

Words from all 45 text passages retrieved by "eliminate" were counted, identifying frequent words that include "competition" and "reduce." It is interesting to note that "reduce" is both a frequent word and a bridge verb in this case.

This method produced 288 total bridge verbs for all search words used, ranging from multiple single occurrences of unique verbs to 122 occurrences of the verb "to be." The verb "defend" occurred 10 times within the overall verb set. Other verbs include those that appear useful e.g., "circulate," "inject" and "eat," as well as common verbs e.g., "have," "be" and "help," that may be less promising as search words. "Have" and "be" occur very frequently as they are used in conjunction with other verbs to form perfect tenses and the passive voice respectively.

Sort dictionary - www.biology-online.org (2005) contained over 65,000 definitions, some of which were not directly related to biology. Some words had biology definitions as well as "regular," more general definitions. Therefore, only definitions tagged as those pertaining to biology e.g., cell biology, microbiology, molecular biology, radiobiology, botany, zoology, ecology, biochemistry and chemistry were extracted. A few subject tags, e.g., mathematics, law and engineering were not as obviously related to biology. However, these types of definitions were retained as they often contribute to the overall understanding of the biological definitions.

Count and rank words within retained definitions – Words from retained definitions were counted and sorted. Those on a stoplist, a list of the most common English words (Leech et al., 2004) were removed from consideration. A word stemmer (Phillipps, 2004) was used to reduce words to their root form, e.g., eliminating, eliminated and elimination have a common root form, eliminate. All of the forms, i.e., eliminate, eliminated, eliminating, elimination etc., contributed to the count of "eliminate."

Determine dictionary definition counts of bridge verbs – The above-described word counts were found for the bridge verbs for "remove." Bridge words were sorted by descending number of occurrences in the dictionary.

RESULTS FOR "REMOVE"

Of the 288 bridge words for "remove," 122 (42.4%) were biologically significant, i.e., contained in terms defined in a biology dictionary (Hine & Martin, 2004; www.biology-online.org, 2004). Figure 2 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 2 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 dark region for the "remove" dataset centers on words that have 70-94 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 significance may be a function of dictionary-definition count, the cumulative density of the biologically significant verbs was plotted against the logarithm of the dictionary definition counts in Figure 3. In this figure, the steepest part of the curve represents the area with the majority of biologically significant words, with the densest region described above approximately located at the inflection point.

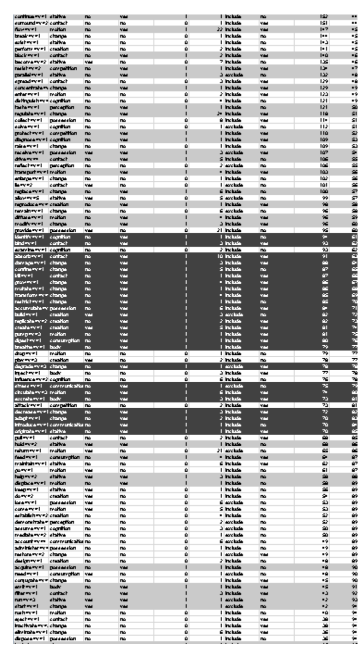

Figure 2: Bridge verbs for "remove" shaded.

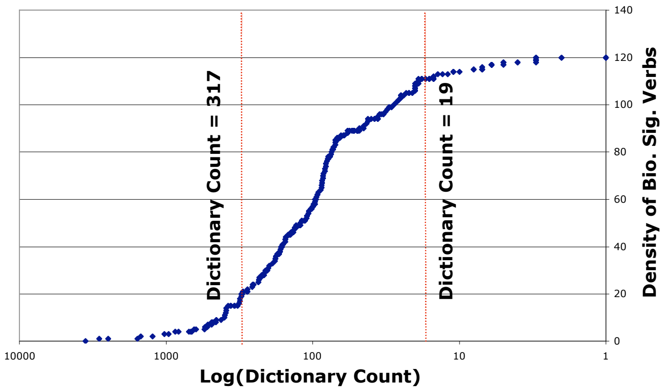

Figure 3: Density distribution graph for "remove."

In Figure 3, 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 biologist, occurred 34 times within the dictionary count, placing it within the range where the majority of biologically significant words occurred. The high concentration of biologically significant words suggests the usefulness of the remaining, biologically connotative, words, e.g., defend, within the boundaries. This method is thus able to identify the non-obvious, but highly relevant keyword suggested by the domain expert in an objective manner.

Other biologically connotative words for "remove" are shown in Table 1.

Table 1: Biologically connotative verbs for "remove" dataset within boundaries.

Word	Dictionary 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

Some words in Table 1 appear obviously related to the original keyword. For instance, "attach" can be seen as an antonym for "remove" even though this relationship is not documented in WordNet or a thesaurus.

RESULTS FOR "ENCAPSULATE"

Thus far, we have described a method that was able to identify a non-obvious but very useful search word suggested by a domain expert. Next, we will confirm that this method can produce useful, biologically connotative words that are not lexically related to the original search words for another example. This next example involves encapsulating pigments to improve stability. Thus biological analogies to encapsulating or enclosing are sought.

"Encapsulate" and "enclose" were the initial search words used to generate the bridge verbs. "Encapsulate" was the original search word, however, it only yielded 2 matches, prompting the use of its hypernym "enclose." "Rupture," a biologically significant word, is an example bridge verb for "encapsulate." The number of times "rupture" occurred in the definitions of the biology dictionary is 39. While "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 yet formalized in lexical references.

For "encapsulate," there were 76 bridge verbs, 31 (40.8%) of which were biologically significant. While 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 4 shows the densest region of biologically significant words for "encapsulate," which fall between 110-120 occurrences in the dictionary. In this region, there are 4 words, all biologically significant.

Figure 5 shows the density distribution plot for the "encapsulate" dataset. It exhibits similar characteristics to the "remove" dataset, 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 26/54 (48.1%) are biologically significant.

Table 2 lists bridge verbs with biological connotations 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 large as the search for "remove," as fewer alternative keywords were generated for "encapsulate." The fact that troponyms/hypernyms appear automatically within the bridge verb dataset suggests that it is not necessary to exhaustively generate alternative search words, or to exhaustively search with them initially.

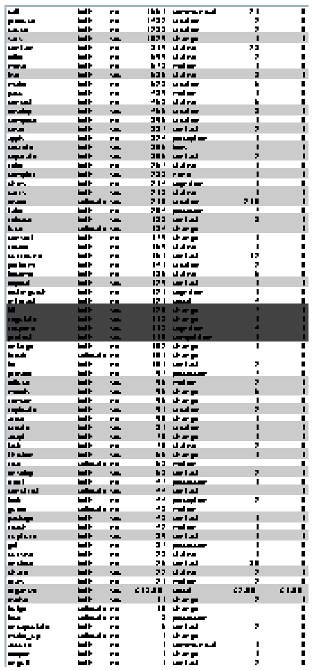

Figure 4: Bridge verbs for "encapsulate."

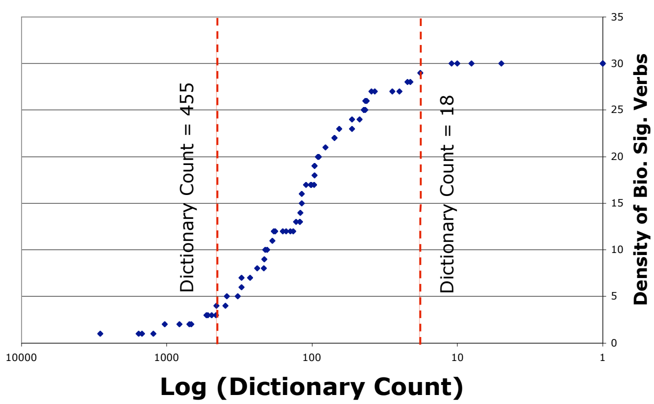

Figure 5: Density distribution graph for "encapsulate."

Table 2: Biologically connotative verbs for "encapsulate" dataset within boundaries.

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

Of the words in Table 2, "survive" probably has the highest intuitive biological connotation. Therefore we will examine some of the matches found in Purves et al. (2001) by searching for forms of "survive."

> 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, i.e., 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 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. Thus, the method described in this paper is able to automatically and objectively identify bridge words such as "survive" and "defend" that link cross-domain lexicons without expert knowledge or reliance on lexical references.

DISCUSSION

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 definitions of biological terms in a biology dictionary. The density of biologically significant words was plotted as a function of dictionary count. These plots show that over 75% (78.7% and 83.9% respectively for the "remove" and the "encapsulate" datasets) of the biologically significant words can be found within boundaries. Both the "remove" and "encapsulate" datasets also contain densest regions of biologically significant words within these boundaries.

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 collocated with the original search words for the clean/remove problem commenced the process of algorithmically identifying "defend," without relying on expert insight, nor on lexical references, which do not list all relationships exhaustively.

Figure 6 below outlines the bridging process.

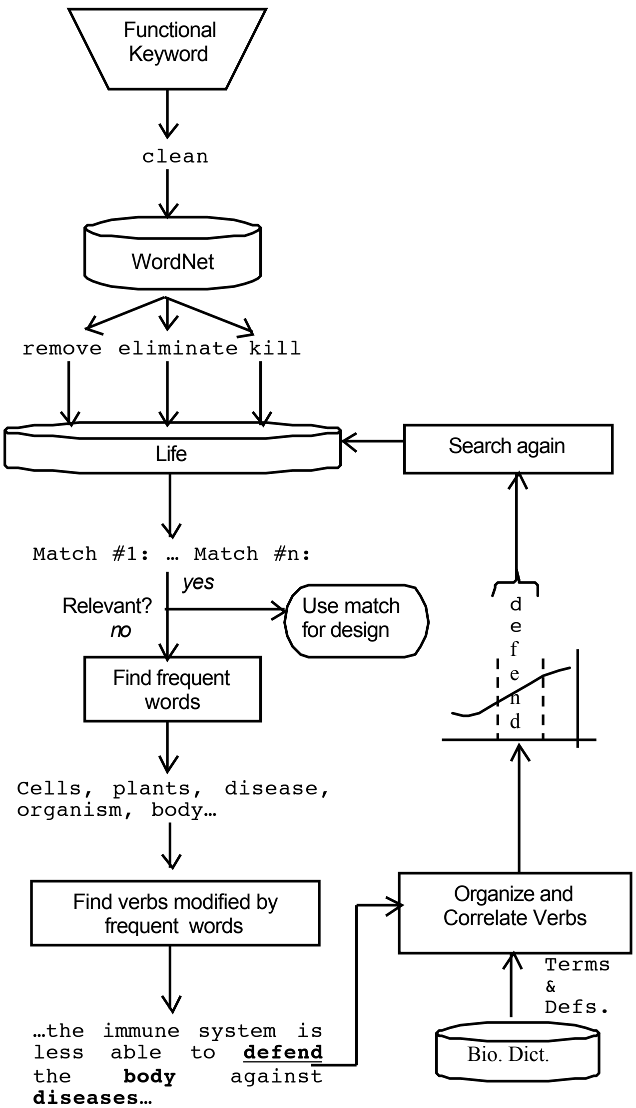

Figure 6: The bridging process.

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, i.e., 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.

From the "encapsulate" dataset, "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 is that "encapsulating" enhances "surviving," just as "removing" enhances "defending." In the case of a plant, it is necessary to sacrifice or remove infected parts to defend itself, just as it is necessary for a cell to encapsulate its DNA to survive adverse conditions.

While 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 20) appear more useful than words that lie above the upper boundary (i.e., above dictionary count 400), as they are less common. Words above the upper boundary may occur too frequently within the biology lexicon to return many meaningful biological phenomena. Interesting words below the lower boundary for the "remove" dataset include "deplete," "reject," and "invade" while words above the upper boundary include "use," "act" and "call." The words above the upper boundary tend to be very 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 dataset include "bulge" and "engulf." The relationship between "bulge" and encapsulate is less obvious, while "engulf" is actually a troponym of "enclose," one of the original search words used.

Two implications regarding lexical references as a result of using this method are:

1. It is not necessary to exhaustively generate troponyms initially to use as search words, as many troponyms are generated using this method.

The set of bridge verbs include troponyms of the original search word, suggesting that it is 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."

2. This method overcomes limitations of lexical references and may identify new relationships between words that are not yet formally documented.

While 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).

Documented lexical relationships depend on the reference chosen. While it is possible to consult several lexical references, this method enables the corpus itself to serve as a guide to the authors' representation of lexical relationships.

While 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.

SUMMARY AND CONCLUDING REMARKS

This work aims to improve creativity and innovation in design by facilitating the use of cross-domain analogies, particularly from biological phenomena, as stimulus for concept generation. Instead of relying on the existence of an enormous database of biological knowledge created specifically to support engineering design, we have chosen to take advantage of the large amount of biological knowledge already in natural-language format, e.g., books, journals, etc. Relevant biological analogies for any given design problem have been located by searching for instances of functional keywords and their synonyms, hypernyms, and troponyms, which describe the intended effect of the design solution. Past case studies involving remanufacture and microassembly have demonstrated the viability of this method.

However, the optimal choice of keywords, or search terms, is complicated by the fact that engineers and biologists may use differing domain-specific lexicon to describe related concepts. Therefore, an engineer without sufficient background in biology may not be able to identify keywords with biological connotation that are not lexically related to the engineering keywords.

Our method involved generating bridge words that frequently collocate with the searched keywords. The biological meaningfulness of these bridge words is determined by how frequently they occur within definitions of biological terms in a biology dictionary. This method was able to produce a non-obvious search word suggested by a domain expert for one example in an objective manner. In another example, a useful search word with biological but no lexical relationships to the original keyword was also produced.

Our approach of generating bridge words with biological meaningfulness is generic and can be used to bridge any disparate domains (e.g., engineering and economics). This process can also be highly automated, enabling designers to quickly access relevant concepts in different domains to inspire innovative design.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support of NSERC (Natural Sciences and Engineering Research Council of Canada). The authors also acknowledge the generosity of www.biology-online.org and Purves et al. for providing machine-readable documents.

REFERENCES

Abney, S., 2002, SCOL version 1h, www.vinartus.net/spa/.

Banerjee, S., Pedersen, T., 2003, The Design, Implementation and Use of the Ngram Statistics Package, http://www.d.umn.edu/~tpederse/nsp.html.

Bejan, A., 2000, From Heat Transfer Principles to Shape and Structure in Nature: Constructural Theory, *Transactions of the ASME,* 122:430-449.

Benami, O, Jin, Y., 2002, Creative Stimulation In Conceptual Design. *Proceedings of ASME DETC/CIE*, Montreal, QC, Canada, DETC2002/DTM-34023

Biology Online, 2004, www.biology-online.org, accessed November 2004.

Biology Online, 2005, www.biology-online.org, contents provided January 2005.

Brill, E., 1994, Rule-based part-of-speech tagger, www.cs.jhu.edu/~brill/.

Chiu, I., Shu, L.H., 2004, Natural Language Analysis for Biomimetic Design, *Proceedings of ASME DETC/CIE,* Salt Lake City, UT, DETC2004/DETC-57250.

Deerwester, S., Dumais, S. T., Harshman, R., 1990, Indexing by Latent Semantic Analysis, *Journal of the Society for Information Science, 41(6)* pp. 391-407.

Fellbaum, C., 1993, English Verbs as a Semantic Net, *Five Papers on WordNet,* pp 40-61. ftp://ftp.cogsci.princeton.edu/pub/wordnet/5papers.ps.

Gordon, W. J. J., 1961, *Synectics, The Development of Creative Capacity*, Harper & Row Publishers, New York, NY.

Hacco, E., Shu, L.H., 2002, Biomimetic Concept Generation Applied to Design for Remanufacture, *Proceedings of ASME DETC/CIE,* Montreal, QC, Canada, DETC2002/DFM-34177.

Hine, R.S., Martin, E. (eds.), 2004, *A Dictionary of Biology*, Oxford University Press.

Hon, K.K.B., Zeiner, 2004, J., Knowledge Brokering for Assisting the Generation of Automotive Product Design, *Annals of the CIRP*, 53/1:159-162.

Joanis, E., Stevenson, S., 2003, A General Feature Space for Automatic Verb Classification, *Proceedings of the 10th Conference of the European ACL,* pp. 163-170.

Leech, G., Rayson, P., Wilson, A., 2001, Companion Website for: Word Frequencies in Written and Spoken English: based on British Nat. Corpus, www.comp.lancs.ac.uk/ucrel/bncfreq/.

Li, X., Schmidt, L., 2000, Grammar-Based Designer Assistance For Epicyclic Gear Trains, *Proceedings of ASME DETC/CIE*, Baltimore, MA, DETC2000/DTM-14574.

Lindemann, U., Gramann, J., 2004, Engineering Design Using Biological Principles, *Proceedings of International Design Conf. – Design 2004*, Dubrovnik 5/18-21.

Luhn, H.P., 1959, Auto-Encoding of Documents for Information Retrieval Systems, *Modern Trends in Documentation*, Pergamon Press, London, pp. 45-58.

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, *Proceedings of ASME DETC/CIE,* Salt Lake City, UT, DETC2004/DETC-57303.

Manser, M. (ed.), 2004, The Chambers Thesaurus, Chambers Harrap Publishers Ltd, Edinburgh, UK.

Matthews, P., 1997, *The Concise Oxford Dictionary of Linguistics*, Oxford University Press.

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, ftp://ftp.cogsci.princeton.edu/pub/wordnet/5papers.ps.

Phillipps, I., 2004, Text::English – Porter's stemming algorithm, http:://cpan.org.

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

Rebholz-Schuhmann, D., Kirsch, H., Couto, F., 2005, Facts from Text – Is Text Mining Ready to Deliver? *PLoS Biology*, 65:188-191

Shu, L.H., Lenau, T., Hansen, H., Alting, L., 2003, Biomimetics Applied to Centering in Microassembly, *Annals of the CIRP*, 52/1:101-104.

Sridharan, P., Campbell, M.I., 2004, A Grammar For Function Structures, *Proceedings of ASME DETC/CIE,* Salt Lake City, UT, DETC2004-57130.

Starling, A., Shea, K., 2002, A Clock Grammar: The Use of A Parallel Grammar in Performance-Based Mechanical Synthesis, *Proc. ASME DETC/CIE*, Montreal, DETC2002/DTM-34026.

Stone, R.B., Wood, K.L., 1999, Development of a functional basis for design, *Proceedings of ASME DETC/CIE*, Las Vegas, NV, DETC99/DTM-8765.

Ullman, D., 2003, *The Mechanical Design Process, Third Edition*, McGraw-Hill, New York, NY.

Vakili, V., Shu, L.H., 2001, Towards Biomimetic Concept Generation, *Proceedings of ASME DETC/CIE*, Pittsburg, PA. DETC2001/DTM-21715.

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, Coordinator of Health Research, University of Saskatchewan, Personal communication.

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

Wood, W.H., Yang, M.C., Cutkosky, M.R., Agogino, A.M., 1998, "Design Information Retrieval: Improving Access to the Informal Side of Design," *Proceedings of the DETC (DTM),* Atlanta, GA, Sept. 13-16, 1998.

WordNet, 2.0, http://www.cogsci.princeton.edu/~wn/.

Yang, M.C., Cutkosky, M. R., 1999 "Machine Generation of Thesauri: Adapting to Evolving Vocabularies in Design Documentation," *Proc. of ICED*, Munich, Germany, Aug. 24-26, 1999.

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

Zipf, G.K., 1949, *Human Behavior And The Principle of Least Effort; An Introduction To Human Ecology*, Addison-Wesley Press, Cambridge, Ma.