Proceedings of the ASME 2010 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference
IDETC/CIE 2010
August 15-18, 2010, Montreal, Quebec, Canada

DETC2010-29057

SUPPORTING BIOMIMETIC DESIGN BY EMBEDDING METADATA IN NATURAL-LANGUAGE CORPORA

J. Ke **I. Chiu** **J.S. Wallace** **L.H. Shu***

Dept. of Mechanical & Industrial Engineering
University of Toronto
Toronto, Ontario, Canada
*shu@mie.utoronto.ca

ABSTRACT

Biology is a good source of analogies for engineering design. One approach of retrieving biological analogies is to perform keyword searches on natural-language sources such as books, journals, etc. A challenge in retrieving information from natural-language sources is the potential requirement to process a large number of search results. This paper describes how inserting metadata such as part-of-speech, word sense and lexicographical data for each word in a natural-language source can help users identify relevant biological stimuli for biomimetic design. Although this research is still exploratory, initial qualitative observations demonstrate successful identification and separation of biological phenomena relevant to either desired functions or desired qualities. In addition, by incorporating the aforementioned metadata, we can automatically remove search results where search keywords act on abstract nouns or where keywords are used in irrelevant senses. The benefits of embedding metadata are demonstrated through a case study on the redesign of a fuel cell bipolar plate. In this case study, our method can be used to hide 64% of the search results that are unlikely to contain useful biological phenomena, reducing the effort to systematically identify relevant biological analogies.

1 INTRODUCTION

Analogies from conceptually different domains have been observed to result in more creative design solutions (Benami & Jin, 2002). Gordon (1961) noted that the specific domain of biology provides the richest source of direct analogies. Many successful biomimetic designs support the notion that biology is a good source of analogies. Despite the demonstrated usefulness of biological analogies in design, designers are likely limited by their personal knowledge of biology. Linsey et al. (2007) also support the idea that designers require tools and systematic methods to access cross-domain knowledge. Most biomimetic design methods, such as the systematic reverse engineering of biological systems (Wilson & Rosen, 2007), can benefit from easy access of cross-domain knowledge.

Two main approaches exist for biological knowledge retrieval. One approach to support biomimetic design is to create a database of biological phenomena organized by engineering function (Vincent & Mann, 2002; Lindemann & Gramann, 2004). However, the creation of such a database may be both time consuming and resource intensive. The process may also be subject to the compilers' own knowledge and bias. Additionally, the rapid growth of biological knowledge provides further challenges for the updating of such a system (Rebholz-Schuhmann et al., 2005; Spasic et al., 2003).

Another approach to support biomimetic design involves searching for instances of functional keywords in natural-language biology knowledge sources, e.g., books, papers, etc. Matched text excerpts containing keywords are examined for relevant biological phenomena that can be applied to the engineering problem of interest. This method takes advantage of the extensive biological information already existing in natural-language format. A method was developed to use word collocation and frequency analyses to identify biologically meaningful keywords that bridge the different lexicons of the fields of biology and engineering (Chiu & Shu, 2007). Cheong et al. (2008) translated terms of the Functional Basis into biologically meaningful keywords, which were not obviously related to the functional keywords, to use as search keywords.

1.1 Natural Language Processing

Along with advantages, there are challenges involved with identifying biological analogies in natural-language knowledge sources. One challenge is the potentially high number of matches, some of which are irrelevant. Hacco and Shu (2002), Chiu and Shu (2004), and Shu (2006) described the following situations that may lead to irrelevant matches.

1.1.1 Word sense and part-of-speech

Words may be polysemous, or have multiple senses or meanings. For example, the word "seal" has many senses, ranging from to close off or isolate, to the aquatic mammal. Matches for "seal" in the aquatic mammal sense are not relevant to searches for "seal" in the closing off or isolating sense. Part-of-speech (POS) may help distinguish between some but not all senses. For example, the aquatic mammal sense only occurs in the noun form of "seal", but the closing off or isolating sense may occur in multiple POS, i.e., noun or verb, forms of "seal". Word sense disambiguation (WSD) aims to distinguish between possible alternative interpretations of words. We can clarify the intended search by specifying the sense for polysemous keywords such as "seal".

1.1.2 Abstract objects

Many engineering design problems involve physical objects. A match where the grammatical object of a search keyword is an abstract entity tends not to be as useful. For example, when searching for "stimulate", matches regarding *stimulating* muscles may be more useful than those regarding *stimulating* interest.

1.1.3 Cross-domain lexical differences

Another challenge arises because engineers and biologists use different lexicons or terminology to describe related phenomena. For example, searching for the keyword "clean" for a problem involving cleaning did not locate many useful matches. When a biological domain expert was asked for alternative keywords, the keyword, "defend" was suggested since some organisms clean as a defensive mechanism. However, "defend" is related to "clean" neither intuitively, for most engineers, nor lexically, e.g., as a synonym. Chiu and Shu (2007) developed a bridging mechanism based on certain semantic relations that identifies such non-obvious, but highly relevant, biologically connotative keywords.

Next, we present nomenclature used in this paper before discussing our approach to increase the relevance of our biomimetic search results. We then present and discuss a case study before summarizing and concluding our work.

2 NOMENCLATURE

2.1 Search Nomenclature

Abstract Noun: Abstract, as opposed to physical, entities, e.g., hope and justice versus vehicles and buildings.

Data Mining: Process of extracting patterns from data.

Corpus (plural: Corpora): A large structured set of text. The corpus on which the biomimetic search tool performs is *Life, the Science of Biology* (Purves et al., 2001).

Lexicographic: Semantic relationships between words of a language, e.g., synonymous and antonymous relationships.

Match Excerpt: One or two sentences surrounding the search keyword in a corpus.

Metadata: Data describing other data, e.g., a word's part-of-speech, sense, etc.

Natural Language: Language spoken or written by humans, as opposed to machine language.

Part-of-speech (POS): Linguistic category of words, such as verbs, nouns or adjectives.

Phrase Group: A group of words functioning as a single entity in a sentence, e.g., *the university student* (noun phrase) or *read the book* (verb phrase).

Polysemy: The existence of multiple meanings for a word or phrase.

Query: The text string a user enters into a search engine to search for related information.

Section: A logically divided group of words in a corpus.

WordNet: A lexical database that organizes words according to their relationships to each other.

Word Sense: One of the possible meanings of a word.

Word Sense Disambiguation (WSD): A process of identifying the intended sense of a word based on the surrounding words.

2.2 Fuel Cell Nomenclature

Proton Exchange Membrane Fuel Cell (PEMFC): A type of fuel cell that operates at a relatively low temperature.

Bipolar Plate: An electrically conductive plate that enables gas flow, electron conduction and water removal in a PEMFC.

Catastrophic Flooding: Excessive liquid water buildup that severely reduces fuel cell performance.

Membrane Electrode Assembly (MEA): An assembly consisting of a proton exchange membrane, electrodes and catalyst in a PEMFC.

3 MOTIVATION

The work in this paper aims to overcome some challenges presented by natural-language processing when searching for biological analogies. We believe that the relevance of biomimetic search results can be improved by clarifying the intended sense of the search keyword.

Several steps are required to automatically extract information from natural-language sources. Identifying POS is important for revealing the linguistic structure of large text corpora (Cutting et al, 1992). POS information facilitates higher-level analysis, such as phrase recognition, WSD and search word context recognition. POS identification facilitates processing and comprehension of biomimetic search results, and thus plays a central role in the work described in this paper.

4 OUR PREVIOUS AND CURRENT WORK

A biomimetic search tool was previously developed to search biological knowledge in natural-language format by finding occurrences of keywords describing the engineering problem. While difficulties common to natural-language processing occurred, this approach does not require the tremendous and somewhat subjective task of categorizing all biological phenomena by engineering function. Thus, this natural-language processing approach can readily take advantage of the enormous amount of biological knowledge already in natural-language format. In the current work, we

incorporate a POS, phrase group, and word sense identification system into the biomimetic search tool.

5 IMPLEMENTATION

An objective of this work is to implement automatic POS recognition for each word of the search results produced by the existing biomimetic search tool. Figure 1 shows an overview of the information flow for the biomimetic search process.

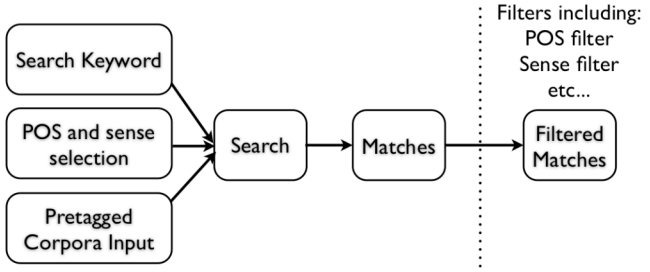

Figure 1. Biomimetic search-tool information flow.

This implementation has three stages of information flow: User input, search of preprocessed corpora, and result output.

5.1 User Input

The user enters a word or a group of words to initiate a search in the biological corpora. In our previous work, verbs were selected as search keywords. Verbs possess several properties desirable for use as design stimuli and best represent engineering functions. We now explore the use of adjectives as search keywords. An adjective designates a particular quality, such as color or shape, to describe a noun. We hypothesized that adjective-based search results will differ from verb-based search results as follows: Verbs can locate new biological phenomena that perform desired functions, while adjectives can locate new biological phenomena that have desired qualities. Thus, the results from the two types of keywords should complement each other.

Our current work also explores the use of other embedded metadata such as word sense in addition to POS. The search tool allows the user to include or exclude senses and POS, as shown in Figure 2, for user input "seal".

SYNSET ID	POS	DEFINITION	NUMBER OF OCCURRENCES	INCLUDE
104160036	n	fastener consisting of a resinous composition that is plastic when warm; used for sealing documents and parcels and letters	30	☐
201354006	v	make tight; secure against leakage	13	☐
100830099	n	the act of treating something to make it repel water	5	☐

Figure 2. Screenshot showing POS and sense selection for user query "seal".

The following example illustrates the ability to search for desired functions versus qualities that can be distinguished using POS. "Thin" has different senses most associated with different POS. When the verb form of "thin" is searched, results describe functions such the "thinning" of bone by a cast that reduces the load on injured bone (Purves et al., 2001). In contrast, when the adjective form of "thin" is searched, results describe how cuticles are thin layers of living tissue that allow the exchange of gases, minerals, and water (Purves et al., 2001).

5.2 Pre-search Calculations

To insert necessary metadata into the corpora, a series of pre-search calculations performed include POS tagging of each word, determining phrase groups within a sentence, identifying abstract nouns and WSD data. This pre-search process needs only be performed once per corpus.

5.2.1 Identifying part-of-speech and phrase group

The Stanford Tagger v1.6 is used to determine the POS and phrase group of each word in the corpora; its algorithm is based on a word's definition and relationship to other words in a sentence (Toutanova et al., 2003). This tagger is chosen as it is over 97% accurate and provides the following features:

a. Can be trained to learn unknown words, i.e., learning algorithms, useful for domain-specific corpora.

b. Able to determine POS of unknown words based on surrounding words, very useful for domain-specific corpora.

c. Able to recognize phrase groups, or substructures of a sentence (e.g., noun groups, verb groups), also known as chunking.

The accuracy of this tagger can be further improved with additional input into the learning algorithm.

5.2.2 Identifying word sense

Each word in the corpora was processed by the Word Sense Disambiguation program (Michelizzi & Pedersen, 2008). In computational linguistics, WSD aims to identify the sense, or meaning, of a word in a given sentence. The English language contains many polysemous words or phrases. By correctly identifying the sense of words in corpora, we can more accurately find relevant biological analogies for biomimetic design. Specifically, incorporating WSD metadata in natural-language corpora may offer the following advantages:

a. Improved relevance of search results by removing matches with keywords in unintended senses.

b. Identified alternative synonymous terms for use as search keywords. For example, "phagocyte" and "scavenger cell" are synonymous. Knowing this synonymous relationship enables return of matches for both "phagocyte" and "scavenger cell" when searching for one or the other.

c. Categorized search results by tracing the hierarchy of hyponyms/hypernyms, or words with more narrow/general meanings.

d. Separation/removal of abstract results by identifying whether a noun is an abstract or physical entity.

Modern WSD engines are only approximately 60% accurate in precision and recall (SemEval-2007). Therefore, to

reduce the effects of incorrect sense tagging, the biomimetic search engine identifies several likely senses of each word based on word occurrences of a tagged corpus with similar context. The WSD algorithm incorporated in the biomimetic search tool is developed by Pedersen et al. (2005) based on second order co-occurrence vectors of pieces of text (Schütze, 1998).

5.2.3 Identifying abstract nouns in phrase groups

After the phrase group is identified, WordNet is used to determine if the noun in the phrase group is a physical or abstract entity. Nouns that describe intangible entities, such as "theory" or "skill" are abstract nouns. A WordNet word sense hierarchy can be constructed by tracing the hypernyms, more general words, or categories, of each noun. For example, Figure 3 shows the hypernym trace of "phagocyte" identifying it as a physical, not abstract, entity.

Figure 3. Hypernym hierarchy classifying "phagocyte" as a physical entity (WordNet v3.0).

Figure 3 shows "phagocyte" classified as an animate thing, then more generally as a "physical entity", with "entity" as the root hypernym. Most nouns are either physical or abstract.

Performing this pre-search processing on a quad-core computer with 4GB of memory took about 24 hours to insert metadata for a standard biology textbook; about 7 words per second. Again, this processing needs to be performed once.

5.3 Corpora Storage

Corpora are divided into individual words and stored in a MySQL information storage system. Each word occupies a row in this system; 20 columns of metadata describe each word. Information such as the POS, word's location in the corpus, word's relation to other words in the phrase group, and lexicographer information are several important metadata that are stored in the MySQL information storage system.

5.4 Search Results

Figure 4 shows that, as typical with search engines, results are displayed as excerpts. When the user clicks on an excerpt, text containing that excerpt is displayed.

The weight and/or color of text in an excerpt denote additional information. For example, a group of bold words denotes a phrase group. When using verbs as search keywords, words that are structurally related (e.g. subject and object) to the search keyword would be shown in red. Knowledge of word relations and sentence structures enables automated selective filtering of the search results.

5.5 Filters

The biomimetic search tool contains two filters to automatically hide search results that are unlikely to contain relevant biological phenomena. The first filter allows users to choose whether to hide results based on the POS of the search keyword. The second filter allows users to choose whether to hide results with search keywords acting on abstract nouns.

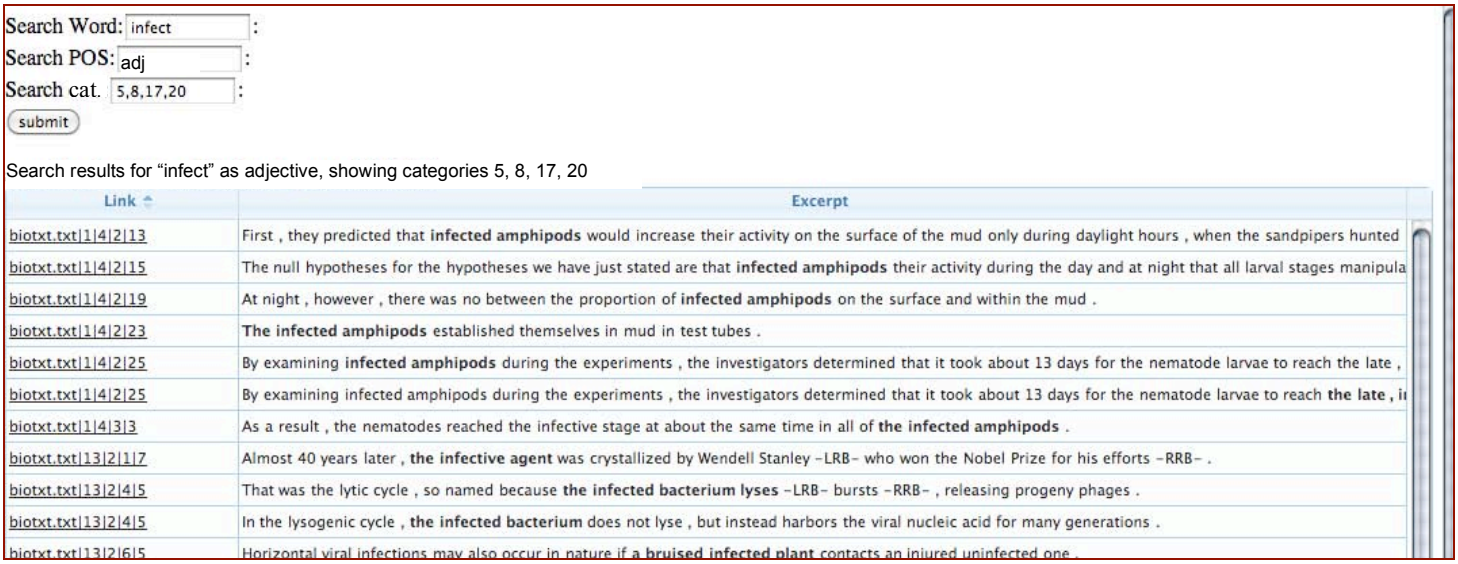

Figure 4. Preliminary interface showing search results for "infect".

6 CASE STUDY

6.1 Problem Statement

Proton exchange membrane fuel cells (PEMFCs) are used in low-temperature power-generation applications. The polymer membrane inside the PEMFC must be properly hydrated and kept at a controlled temperature of ~70 ^{o}C to efficiently conduct protons. Water, a product of the fuel-cell reaction, must also be removed from the cathode side to prevent congestion, or flooding, in the oxidizer flow channels.

The bipolar plate serves as an important component in a PEMFC for water removal and temperature control. Figure 5 illustrates a typical bipolar plate cross-section.

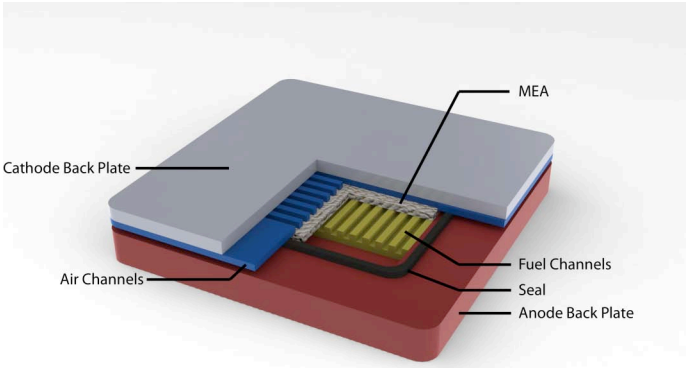

Figure 5. Typical PEMFC bipolar plate configuration (not to scale)

This case study aims to generate concepts to efficiently manage a PEMFC's hydration level. The biomimetic search approach, augmented with newly implemented functionality, is demonstrated and used to identify possible solutions.

6.2 Part-of-speech Selection

Verbs have been predominately used to initiate past biomimetic searches. Past work detail the process of selecting keywords, extracting strategies from phenomena described in result excerpts, and applying the strategies to develop engineering solutions (Chiu and Shu, 2007; Cheong et al., 2009). In fact, one previous paper detailed the authors' attempt to solve the present case study using verbs as search keywords (Ke et al., 2009). In contrast, the present paper approaches the case study by using adjectives as search keywords.

6.3 Source of Biological Information

The initial source of biological information, *Life, the Science of Biology*, by Purves et al. (2001), hereinafter referred to as the corpus, is a reference text for an introductory biology course at the University of Toronto. This text is suitable for the purposes of this approach for two reasons. First, the book is written at a level that is easily understood by those who have little or no background in biology. Second, the book covers several levels of biological organization, from the molecular (e.g., DNA) to the ecosystem level. As previously reported, while this initial source may not give enough details to fully develop a novel solution, it is useful for identifying relevant phenomena that can then be further researched in more advanced sources. Subsequent sources could involve more specific texts to provide details on relevant phenomena.

6.4 Search Keywords

The design objective is to transport water efficiently. The corresponding quality or adjective that accurately describes this design objective was initially unclear to the authors. Therefore, WordNet was used to explore and identify potential search keywords. The initial starting keyword, "wet" was chosen for two reasons. First, "wet" is a common adjective with several related adjectives within the WordNet hierarchy, increasing our chances of finding more relevant search keywords. Second, "wet" loosely describes the design objective, increasing the resulting excerpts' chance of containing biological phenomena that can lead to design concepts.

Our past work found WordNet relationships to be excellent sources for alternative keywords (Hacco & Shu, 2002; Chiu & Shu, 2004). Since antonyms and synonyms are key relationships for adjectives, they were used to obtain additional search keywords. Figure 6 shows the partial exploratory path taken to find search keywords within WordNet.

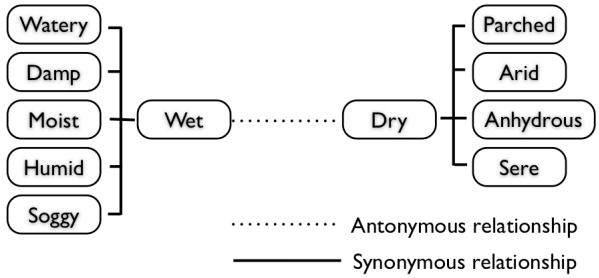

Figure 6. Related words of "wet" based on WordNet v3.0 relationships.

The following adjectives were chosen as final keywords: damp, evaporative, excretory, flooded, fluid, humid, soggy, hydraulic, hydrophilic, moist, osmotic, wet, and dry.

6.5 Results

There were 180 unique result excerpts, of which 28 were automatically removed as the match excerpt used the keyword in a different POS than specified. Some adjectives can also be used as nouns and verbs. For example, while more useful to this case study as an adjective, the word "fluid" occurs very commonly in the corpus as a noun in less relevant contexts. Nouns are generally not preferred as search keywords, as nouns tend to bias designers towards preconceived biological phenomena (Chiu & Shu, 2004) and not to provide analogies with respect to functions or qualities. By identifying the POS of the keyword, the biomimetic search tool can selectively omit search results with search keywords with a particular POS.

Within the remaining 152 result excerpts, 87 have search keywords acting on abstract nouns; therefore, they can be automatically excluded from the results. For example, abstract adjective-noun pairs that were removed include wet-conditions and wet-seasons. Of the 180 original result excerpts, 65 passed

the filtering process, a 64% reduction in the number of excerpts that would require manual processing. Figure 7 shows the POS and sense selection screen for matches of keyword "wet".

SYNSET ID	POS	DEFINITION	NUMBER OF OCCURRENCES
114535431	n	wetness caused by water	18
200214951	v	cause to become wet	2
302547317	a	covered or soaked with a liquid such as water	13

Submit

Figure 7. POS, senses, occurrences of "wet".

After inspecting the remaining 65 matches and performing cursory research on the phenomena described, two promising biological phenomena were selected for further consideration.

a. Water transport in plants based on cohesion-tension

This phenomenon was found using the keyword "evaporative", and is described in the following partial excerpt:

> *Transpiration, the **evaporative** loss of water from the leaves, generates a pulling force (tension) on the water in the apoplast of the leaves* (Purves et al., 2001).

We chose to investigate this particular phenomenon due to similar water transport difficulties faced by leaves and fuel cell bipolar plates. Specifically, both leaves and fuel cell bipolar plates require a mechanism to move water through small channels with relatively high flux.

In cohesion-tension theory, which was first proposed by H.H. Dixon (1914), water is drawn up in plants by transpiration. Due to the cohesive and adhesive properties of water, as water molecules evaporate from the stoma, additional water molecules are pulled from the lower part of the plant through xylem, a type of transport tissue.

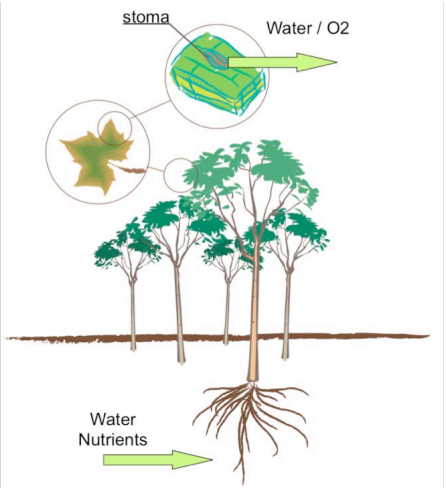

Figure 8. Water transport in plants

b. Water transport based on capillary action

This phenomenon was found by searching for the keyword "damp", and was described in the context of the below match:

> *They (nontracheophytes) often are found on **damp**, cool ground, where they form thick mats* (Purves et al., 2001).

We chose to investigate this particular phenomenon due to the similarities between the fuel cell operating environment and the water transport objective of nontracheophytes. Specifically, both nontracheophytes and fuel cell bipolar plates operate in damp environments and require a mechanism to transport moisture through a barrier.

The text section corresponding to the above match explains that nontracheophytes (liverworts, hornworts, and some mosses) usually inhabit moist environments but lack vascular tissue to circulate liquid (Purves et al., 2001). Nontracheophytes contain several physical features that assist external water conduction. For example, overlapping leaves, rhizoids (shown in Figure 9), ridged leaves, and tiny warts help water conduction by maximizing the effectiveness of capillary action.

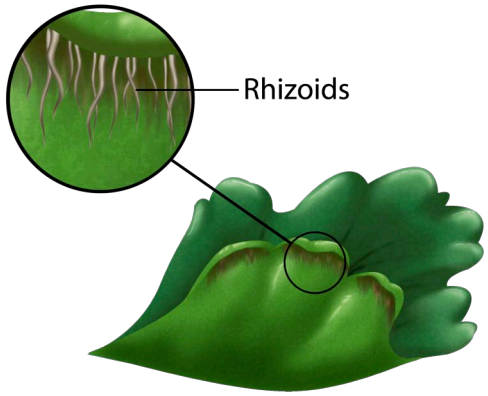

Figure 9. Rhizoids on liverworts

6.6 Generated Concept

A concept was generated with help from the members of the biomimetic fuel-cell design group of the Biomimetics for Innovation and Design Laboratory at the University of Toronto. The proposed redesign, shown in Figure 10, contains two main components to facilitate water removal:

a. Polyethersulfone (PES) hollow-fibers

Similar to xylem in plants, PES hollow-fibers must be continuous and under tension to maximize water flux.

b. Machined micro-channels

Similar to ridged leaves, micro-channels maximize capillary action and direct excess water to the PES hollow-fiber for water removal.

Figure 11 shows the analogical mapping of the biological phenomena to the solution in the engineering domain.

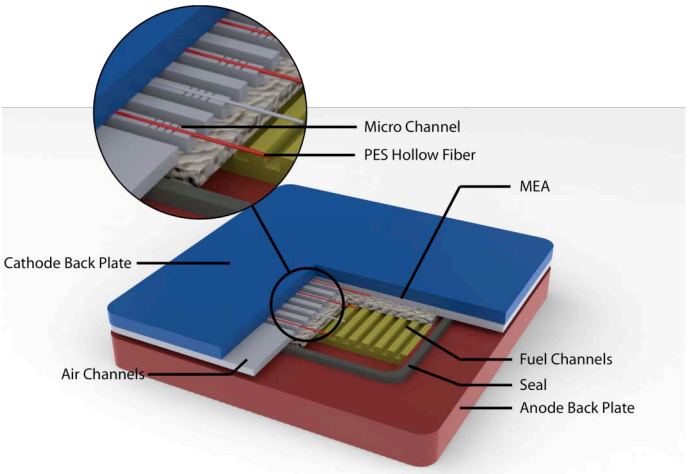

Figure 10. Redesigned PEMFC bipolar plate configuration (not to scale).

Biological Domain		Engineering Domain
Xylem and water transport in plants	>	PES hollow tubes for water removal
Ridged leaves	>	Machined micro-channels to induce capillary action

Figure 11. Mapping from biological phenomena to the redesigned PEMFC bipolar plate.

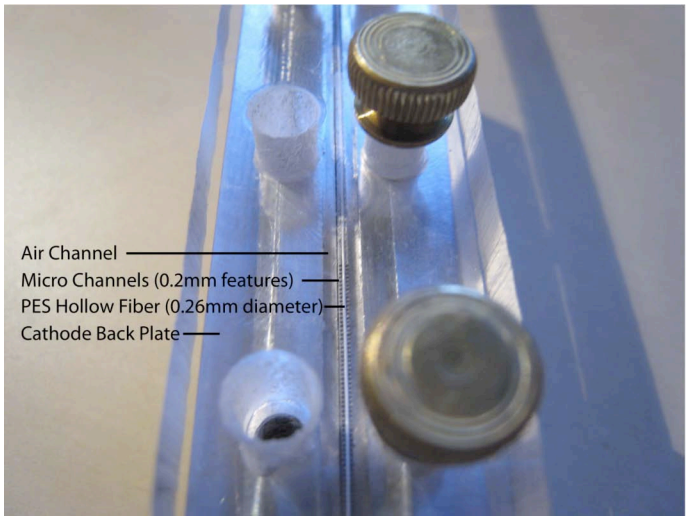

Figure 12. Redesigned bipolar plate prototype on test rig.

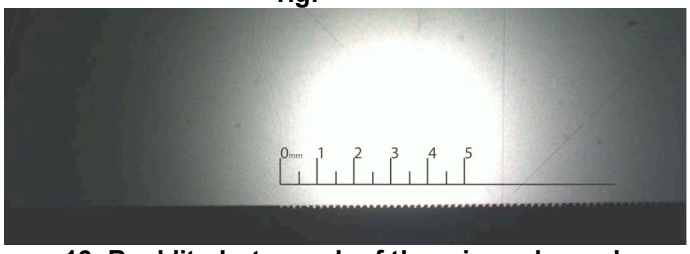

Figure 13. Backlit photograph of the micro-channel features that induce capillary action.

Several prototypes, one shown in Figures 12 and 13, were created to validate the proposed concept. The prototypes were demonstrated on a test apparatus with low airflow rates and low inlet air pressure. Preliminary experiments result in total recovery from flow channel catastrophic flooding within seconds.

7 DISCUSSION

The purpose of implementing POS, phrase group and word sense identification is to address two challenges associated with natural-language search for biological phenomena. The first challenge is that a large number of results may exist for a particular search. Therefore, the user may need to process a large amount of information to identify the most suitable analogies. The second challenge is that not all results are relevant; for example, irrelevant results may include keywords with an undesired POS, incorrect word sense, or keywords acting on abstract nouns. Identifying the word POS, sense and phrase groups in the corpora makes possible the following:

1. Separate results by search keyword POS and sense

Search keywords in the verb POS may locate biological phenomena that perform related functions. In contrast, search keywords in the adjective POS may locate biological phenomena relevant to desired qualities. After specifying the intended search word POS, users can further narrow down search results by showing only result excerpts that also use the intended search word sense.

2. Remove results where keywords act on abstract nouns

Since mechanical design usually involves physical objects, we can also choose to remove matches that contain search keywords acting on abstract nouns. The introduction of POS, word sense and phrase group identification enables this task to be done automatically by identifying grammatical objects that are abstract entities.

The preceding implementation addressed some of the challenges of using natural-language corpora to search for biological phenomena. As the present research is still exploratory, future work includes validation of the effectiveness of our implementation.

8 SUMMARY AND CONCLUDING REMARKS

This paper describes the process and advantages of attaching information such as POS, word sense, phrase group and consequent identification of abstract/physical objects to each word of text to be searched. By inserting metadata into the corpora, the search results can be better organized for more efficient review.

Natural-language search for biological phenomena is a powerful approach to find stimuli for biomimetic design despite some challenges. The presented method attempts to address two of the challenges: A large number of matches and possible irrelevant matches.

Initial efforts reveal a positive outlook for the present approach. Qualitatively, by choosing to limit results to those

that contain matches to keywords in specific POS, users can choose to display results related to desired functions or desired qualities. Also, tagging phrase groups with WordNet lexicographer metadata enables users to omit search results that have search keywords acting on abstract versus physical nouns.

A redesign of a water management system for the PEM fuel cell serves to illustrate this approach. In this case study, the benefits realized of the embedded metadata include reduction in the effort required to identify several relevant biological phenomena from a large number of search results. Selected biological phenomena were used to generate a design concept, which was subsequently prototyped with promising results.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support of the Natural Sciences Engineering Research Council of Canada through a Strategic Projects Grant: "Biomimetic design applied to fuel cell redesign." Thanks are due to Jessica Currie and Keith Fung for helpful discussions on fuel cell concepts.

REFERENCES

Benami, O, Y. Jin, 2002, Creative Stimulation in Conceptual Design, *Proc. ASME DETC/CIE,* Montreal, DETC2002-DTM-34023.

Cheong, H., L.H. Shu, R.B. Stone, D.A. McAdams, 2008, Translating Terms of the Functional Basis into Biologically Meaningful Keywords, *Proc. IDETC/CIE*, New York City, NY, DETC2008-49363.

Cheong, H., L.H. Shu, 2009, Effective Analogical Transfer Using Biological Descriptions Retrieved with Functional and Biologically Meaningful Keywords, *Proc. IDETC/CIE,* San Diego, CA, DETC2009/DTM-86681.

Chiu, I., L.H. Shu, 2004, Natural Language Analysis for Biomimetic Design, *Proc. IDETC/CIE,* Salt Lake City, Utah, DETC2004/DTM-57250.

Chiu, I., L.H. Shu, 2007, Biomimetic Design through Natural Language Analysis to Facilitate Cross-Domain Information Retrieval, *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 21/1:45-59.

Chiu, I., L.H. Shu, 2008, Effects of Dichotomous Lexical Stimuli in Concept Generation, *Proc. IDETC/CIE,* New York City, NY, DETC2008/DTM-49372.

Cutting, D., J. Kupiec, J. Pedersen, P. Sibun, 1992, A Practical Part-of-speech tagger, *Proc. Third Conference on Applied Natural Language Processing,* Trento, Italy.

Dixon, H.H., 1914, *Transpiration and the Ascent of Sap in Plants,* Macmillan and Company, London, England.

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

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

Ke, J., J.S. Wallace, L.H. Shu, 2009, Supporting Biomimetic Design Through Categorization of Natural-Language Keyword-Search Results, *Proc. IDETC/CIE,* San Diego, CA, DETC2009/DTM-86681.

Lindemann, U., J. Gramann, 2004, Engineering Design Using Biological Principles, *Proc. Int. Design Conf.* Dubrovnik, Vol. 5 pp. 18-21.

Linsey, J.S., J.P. Laux, E. Clauss, K.L. Wood, A.B. Markman, 2007, Increasing Innovation: A Trilogy of Experiments towards a Design-by-analogy Method. *Proc. ASME IDETC/CIE*, Las Vegas, NV, DETC2007-34948.

Michelizzi, J., T. Pedersen, WordNet::SenseRelate::AllWords, 2008, http://search.cpan.org/~tpederse/WordNet-SenseRelate-AllWords-0.19/lib/WordNet/SenseRelate/AllWords.pm (last accessed May 3, 2010).

Pedersen, T., S. Patwardhan, S. Banerjee, 2005, WordNet::Similarity::vector, http://search.cpan.org/~tpederse/WordNet-Similarity-2.05/lib/WordNet/Similarity/vector.pm (last accessed May 3, 2010).

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

Rebholz-Schuhmann, D., H. Kirsch, F. Couto, 2005, Facts from Text - Is Text Mining Ready to Deliver, *PLoS Biology*, e65.

Schütze H., 1998, Automatic Word Sense Discrimination, *Computational Linguistics*, 24(1), pp. 97-123.

SemEval-2007 web page, http://nlp.cs.swarthmore.edu/semeval (last accessed May 3, 2010).

Shu, L.H., 2006, Using Biological Analogies for Engineering Problem Solving and Design, *Proc. CDEN/RCCI,* Toronto, Canada.

Spasic, I., G. Nenadic, S. Ananiadou, 2003, Using Domain-specific Verbs for Term Classification, *Proc. ACL Workshop on Natural Language Processing in Biomedicine,* Sapporo, Japan, pp. 17-24.

Toutanova, K., D. Klein, C.D. Manning, Y. Singer, 2003, Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network, *Proc. HLT-NAACL,* Edmonton, Canada.

Vakili, V., L.H. Shu, 2001, Towards Biomimetic Concept Generation, *Proc. ASME IDETC/CIE*, Pittsburgh, PA, DETC2001/DTM-21715.

Vincent, J., D. Mann, 2002, Systematic Technology Transfer from Biology to Engineering, *Philosophical Transactions of The Royal Society: Physical Sciences*, 360:159 - 173.

Wilson, J.O., D.W. Rosen, 2007, Systematic Reverse Engineering of Biological Systems. *Proc. ASME IDETC/CIE*, Las Vegas, NV, DETC2007-35395.

WordNet – Princeton University Cognitive Science Laboratory. 2010. http://wordnet.princeton.edu (last accessed February 3, 2010).

