Proceedings of the ASME 2009 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference
IDETC/CIE 2009
August 30 - September 2, 2009, San Diego, California, USA

DETC2009-86681

SUPPORTING BIOMIMETIC DESIGN THROUGH CATEGORIZATION OF NATURAL-LANGUAGE KEYWORD-SEARCH RESULTS

Ji Ke **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 of retrieving information from natural-language sources is the potential requirement to process a large number of search results. This paper describes a categorization method that organizes a large group of diverse biological information into meaningful categories. The benefits of the categorization functionality are demonstrated through a case study on the redesign of a fuel cell bipolar plate. In this case study, our categorization method reduced the effort to systematically identify biological phenomena by up to ~80%.

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 that designers require tools and systematic methods to access 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 resource 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 biology knowledge sources in natural-language format, 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 in the fields of biology and engineering (Chiu & Shu, 2007). Cheong et al. (2008) translated terms of the Functional Basis into biologically meaningful keywords, not obviously related to the functional keywords, to use as search keywords.

However, a shortcoming of searching natural-language text using keywords is the large number of matches that may result. Therefore, the designer may need to process a large amount of information to identify the most suitable analogies.

NOMENCLATURE

Search Related

Data Mining: Process of extracting patterns from a collection of 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).
Match Excerpt: One or two sentences surrounding the search keyword in a section.
Section: A logically divided group of words from the corpus.
WordNet: A lexical database that organizes words according to their relationships to each other (WordNet, 2006).
Stop Words: A list of commonly occurring words disregarded during searches, e.g., the, for, etc.

Super-category: A category from which other categories are derived.
Sub-category: A category that is derived from its parent category.
Category: Sub-categories and sections that share a common idea implied by the category name.
Category Trigger: A word or a group of words that results in a section to be classified by a category.
Metadata: Data describing other data.

Fuel Cell Related

PEMFC (Proton Exchange Membrane Fuel Cell): a type of fuel cell that operates in relatively low temperatures.
Bipolar Plate: A plate (usually electrically conductive) that typically enables gas flow, electron conduction and water removal in a PEMFC.
Membrane Electrode Assembly (MEA): An assembly consisting of a proton exchange membrane, electrodes and catalyst in a PEMFC.

MOTIVATION AND APPROACH

The work described in this paper aims to overcome the obstacle of processing potentially large numbers of matches when searching for biological analogies to support design. Specifically, a method to more effectively categorize and display search results or matches is developed and presented. We believe that this work will facilitate processing and comprehension of biomimetic search results, thus supporting the use of biological analogies in design. Biomimetic design methods such as the systematic reverse engineering of biological systems (Wilson & Rosen, 2007) can benefit from the result of this research.

Librarians have long used classification systems to enable categorization. More recently, some computer search engines have directory systems to help cope with the exponential growth in the number of web pages. In the design domain, a number of ontologies (Stone & Wood, 2000; Kitamura et al., 2006; Gero & Kannengiesser, 2007) have been developed to structure and describe design concepts and their relationships. However, these categorization systems still require people to manually identify categories and perform the categorization. To manually categorize all of biological information for the purpose of biomimetic design is a daunting task. This paper develops and describes an automated categorization method.

Two basic principles exist for categorization (Levitin 2002). The first states that categorization should provide maximum information with the least cognitive effort. The second principle states that categorization information must be in the form of structured entities. Both principles are applied in the creation of the present computerized categorization method.

This automated categorization method provides the following benefits. Sorting a large number of search results into categories facilitates information navigation. The categories provide an overview in the form of one-to-two word summaries of the matched text excerpts. This overview quickly presents the diversity of the biological phenomena associated with the search keyword, several of which may reveal unexpected solutions. Upon surveying the category contents, the designer can more quickly focus on those matches that are most promising, as well as filter out less promising results.

BACKGROUND

Search result categorization is actively researched for many diverse applications, e.g., search engines, library cataloguing, management, linguistics, data mining, etc. Most categorization techniques require the user to manually assign a category to a database entry before a category search can be performed. For example, the ontology-based annotation framework, a metadata scheme that uses functionality based semantics to relate information content of engineering design documents requires the authors to insert the metadata into the design documents (Kitamura et al., 2006).

Currently, challenges exist in using computational language processing techniques to automatically categorize without human intervention. Perhaps this is the reason some major search engines still do not have an automatic categorization system in place. Nevertheless, some approaches toward real-time categorization follow.

Hao et al. (2008) use name-entity recognition and part-of-speech analysis to retrieve a verb-noun pair from a search result. They then use a semantic similarity algorithm based on WordNet to calculate the similarity between each text excerpt and each pre-defined category to determine the category for the match. A preliminary experiment using this method demonstrated that 72.7% of the matches are correctly categorized.

Pedersen et al. (2007) used six existing domain-independent measurements to compute the similarity between information in a biomedical context. The semantic measurements include three path-based ones using the WordNet hierarchy, two information content-based measurements, and a vector-based measurement. This last measurement creates context vectors that represent the meaning of concepts derived from word frequency analysis of several corpora. However, when compared with the results of manual similarity assessment by nine medical professionals, the results of these six computational methods of similarity measurement were generally poor.

Chen and Dumais (2000) describe the Support Vector Machine (SVM) that used 13,352 pre-classified web pages to train a model. In practice, fewer than 70% of the webpages were correctly identified by the SVM method. However, test subjects were observed to find information 50% faster once the websites have been categorized.

Schunn and Vera (1995) experimentally determined the factors that contribute to category structure. These factors include linguistic and cognitive properties such as casual/functional relevance to a category, frequency of association with a category, and the value of the property for recognizing a member of the category.

OUR PREVIOUS WORK AND CURRENT APPROACH

Our current categorization approach is integrated into a biomimetic search tool previously developed to locate in biological knowledge in natural-language format, 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 approach can readily take advantage of the enormous amount of biological knowledge already in natural-language format.

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 the 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 at a level that is easily understood by those who have little or no background in biology. Second, the book is general and covers several levels of biological organization, from the molecular (e.g., DNA) to the ecosystem level. A search tool was developed that looks for occurrences of functional words describing the engineering problem within the electronic copy of this text. As previously reported, while this initial source may not give enough details to inspire 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 find details on relevant phenomena such as those on molecular and cell biology, plant physiology and animal physiology.

IMPLEMENTATION

The specific objective of this work is to implement a computerized method that categorizes search results from the existing biomimetic search tool. The corpus is divided into sections and stored in a MySQL database that contains 1435 sections, averaging 365 words per section. Figure 1 shows an overview of information flow for the categorization process:

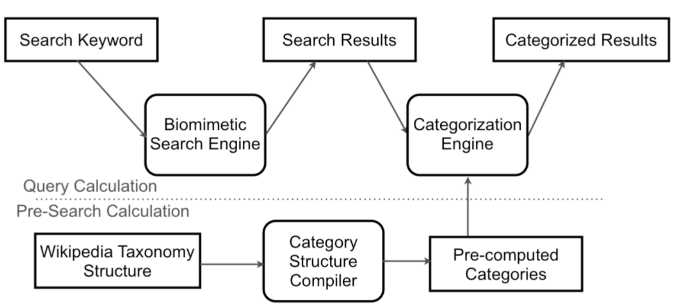

Figure 1 – Overview of Information Flow between Biomimetic Search Tool and Categorization Tool

Pre-Search Calculation

Wikipedia Taxonomy Structure

Before the actual categorization of search results, a pre-search calculation is used to determine the categories; this process needs only be performed once. The pre-search calculation extracts category structures from the "Biology" category entry from Wikipedia. Wikipedia's content has been used extensively for academic research (Prescott, 2006). Wikipedia categories have recently been used to identify concepts (Syed et al., 2008). Furthermore, the biological ontology structure within Wikipedia is more detailed, better structured, and better referenced than the ontologies of other promising resources such as Bio-Medicine.org (Bio-Medicine, 2009) and Bio-Online.org (Bio-Online, 2009).

Both categories and articles are extracted from Wikipedia. Categories consist of super-categories and sub-categories. "Biology" is the only super-category for the current biomimetic search tool; other super-categories will be considered for future work. The resulting taxonomy structure is predominantly hierarchical, with some redundant and possibly undesired interconnections, an example of which is shown in Figure 2.

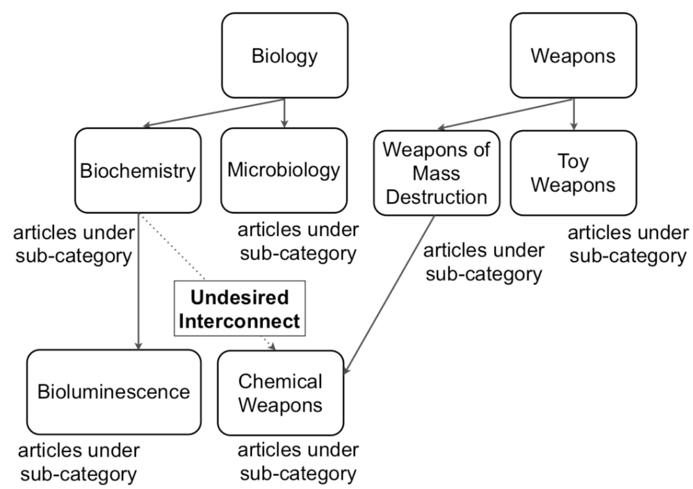

Figure 2 – Representation of Wikipedia-Based Taxonomy Structure

Category Structure Compilation

A computer program extracts categories and corresponding articles from Wikipedia. The titles of the articles are used as category triggers. A category trigger is a word or a group of words that the categorization engine will search for in each match excerpt. If the trigger is found, then the matched section will be classified under the category that triggered it.

Sub-categories are hierarchical and exist at different levels. Immediately direct sub-categories of the super-category "Biology" are labeled level 2, and subsequent sub-categories are more specific than the previous level, and are numbered incrementally. We observed that higher numbered categories are more likely to contain detailed and technical articles, increasing their value as design stimuli.

However, more specific categories are also more likely to be noise categories. Noise categories are encountered due to the interconnectedness of the Wikipedia taxonomy structure. An example noise category is “Chemical Weapons”, which is a sub-category of both the super categories "Biology" and "Weapons" as shown in Figure 2. Another undesired effect of using a more specific category level is over-categorizing. Over-categorizing arises when categories are too finely defined, which might result in a large number of categories. A category cut-off level is selected to balance the negative effects of noise categories and over-categorizing and the desired qualities of category specificity and useful content.

Figure 3 shows the difference in scope between the different levels of categories. One can choose one or a combination of category levels to view search results. We observed for most problems that category level 4 is a good balance between specificity and usefulness. To enhance search results further, the Porter stemming algorithm is used to remove common morphological and inflexional endings from words (Porter, 2006). A MySQL stop-word list is used to filter out common words that do not add semantic value (MySQL, 2008).

Category Level	Specificity	Sample Category
1	Very General	Biology
2	General	Neuroscience, Genetics, Organisms
3	Specific	Eukaryotes, Bioluminescence, Biomolecules
4	Very Specific	Lipids, Cellular respiration, Neurotransmitters

Figure 3 – Example Categories at Different Levels

Query Calculation

Categorization Engine

To initiate the process, the user enters search keywords. The biomimetic search tool looks for instances of the keywords, generating excerpts consisting of one or two sentences surrounding the search keyword from the corresponding section. The categorization engine then searches for the category triggers within the excerpt. If the category triggers exist within the excerpt, the section that contains the excerpt is categorized under the category trigger.

Categorized Results

After the categorization engine processes the excerpts and sorts them into categories, the results are displayed in three levels, containing increasing detail: 1) the list of categories, 2) the match excerpts, and 3) the sections containing the matches. Figure 4 is a screenshot of example categorized results.

The biomimetic search tool’s user interface mimics the “Column View” mode found in modern operation systems (e.g., Apple’s Finder). In this interface, columns represent different levels of information, increasing in detail from left to right. Column contents are expanded by clicking on a sub-category (represented by a folder icon, Figure 4, label 1) or a section title (represented by a file icon, Figure 4, label 2). The one or two sentences corresponding to the match excerpt appear when the mouse is moved over a section title (Figure 4, label 3). As intended, the category lists give minimal information but provide an overview of the matches. The match excerpt gives more information by displaying one or two sentences surrounding the keyword, and the entire section (Figure 4, label 4) that contains the match excerpt gives the most information.

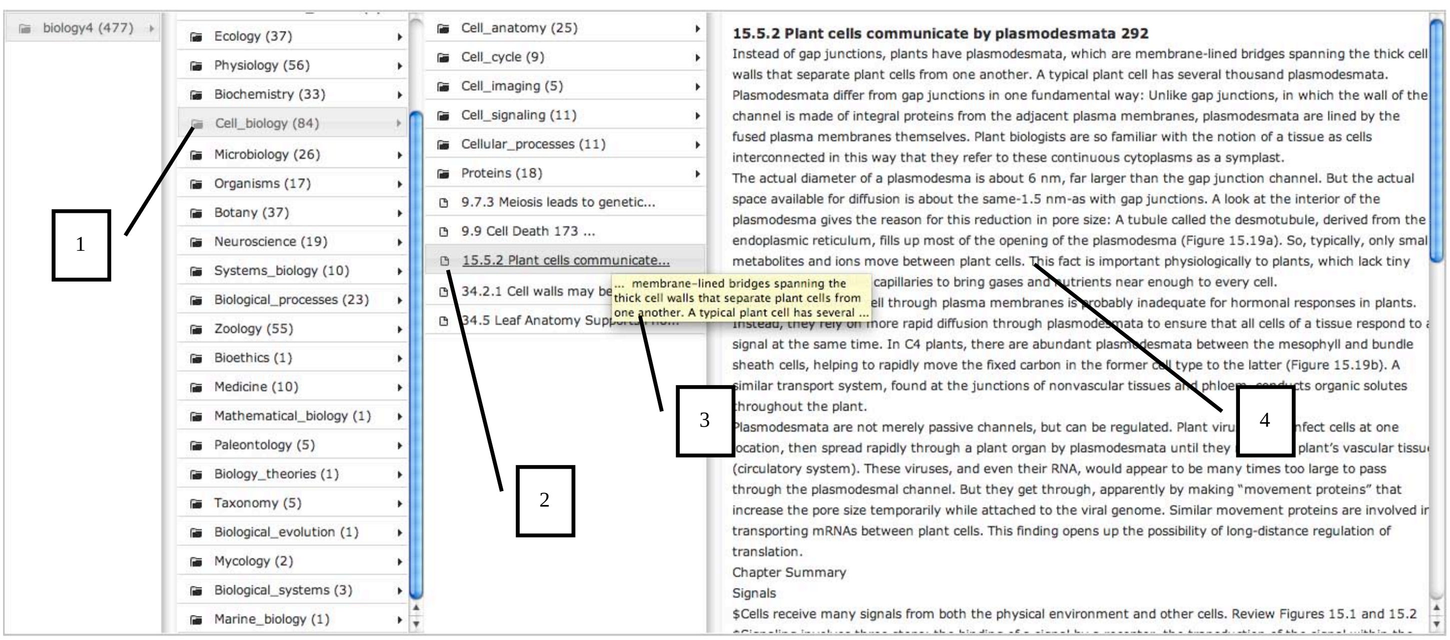

Figure 4 – Screenshot of the Biomimetic Search Tool User Interface

CASE STUDY

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 from 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.

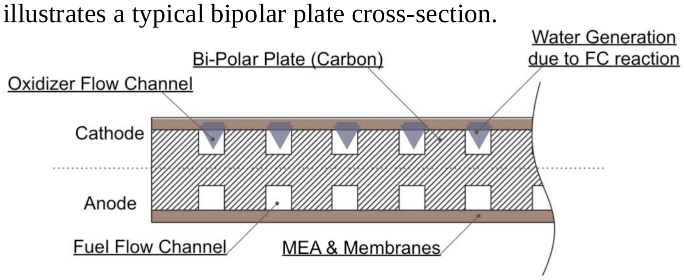

Figure 5 - Typical PEMFC Bipolar Plate Configuration

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

Biomimetic Search

The first step of this process is to select the search keywords. The results of Cheong et al.'s (2008) work that identified biologically meaningful keywords for function terms of the functional basis is used in this case study. The functional basis describes product functions with verb-object pairs in a domain generic format (Stone & Wood, 2000).

The functions Remove + Extract most accurately describe the problem, as water and heat must be removed from the fuel cell. The corresponding biologically meaningful keywords generated by Cheong et al. (2008) include: Splice, Collect, Extract, Trap, Delete, Degrade, Beat, Separate.

After the search results using these keywords are found, categorized, and displayed, the next step is to determine the biological phenomena associated with each category. This can be done in two ways. First, one section excerpt from each category is inspected to assess the content for the rest of the matches in the category. Second, the section content is assessed based on its category name. Either of these methods will reduce the effort required to process the search results.

Results

The biomimetic search tool first presents the search results under level 2 categories (where the single level 1 category in our case is "biology". Level 2 categories such as "physiology" and "organisms" are fairly generic. Higher numbered categories are more detailed, where the category titles themselves may reveal unexpected insights. One such category is "vision", where the match pertains to photosensitivity and the rod cell, rhodopsin. Specifically, the corresponding section explains that the rod cell contains an outer segment of plasma membrane that captures photons of light passing through the cell (Purves et al., 2001). This phenomenon may therefore be relevant to the case study that involves capturing water molecules.

Another interesting level 4 category is "extremophiles." Engineers who are unfamiliar with extremophiles are unlikely to appreciate the relevance of the corresponding matches. The category title may help facilitate comprehension through lexical analysis, glossary definition or a quick search, and reveal that an extremophile is a type of organism that can survive in extreme geochemical conditions. Archaea and euryarchaeota, discussed in the two matches under this category, are types of extremophiles. The matched excerpts highlight the unique cell membrane structures that enable extremophiles to survive in extreme conditions. One section explains that extreme halophiles (a type of euryarchaeota) are able to conserve water and not dry out due to the high salt content of its surroundings. This phenomenon therefore appears relevant for the design of the bipolar plate. Figure 6 shows the sequence of subcategories that led to the above two phenomena.

Biology -> Biological Systems -> Senses -> Vision-> Article regarding rhodopsin

Biology -> Microbiology -> Microbial growth and nutrition -> Extremophiles-> Articles regarding archaea and euryarchaeota

Figure 6 – Categorizing Hierarchy Used to Reach Example Articles

Figure 7 shows example categories for the eight search keywords; Annex A contains a selection of biological phenomena for this case study, derived from 405 matches and 1083 categories.

	Example Level 4 Categories
Splice	Bacteria, Prokaryotes, DNA, Molecular Biology, Proteins
Collect	Infectious Diseases, Parasites, Fossils, Hormones
Extract	Neuroanatomy, Planktology, Toxicology, Mycotoxins
Trap	Cellular Process, Photosynthesis, Cellular Respiration
Delete	Proteins, Applied Genetics, Chromosomes
Degrade	Biomolecules, DNA, RNA, Microbiology
Beat	Body fluids, Eukaryotes, Prokaryotes, Cell Anatomy
Separate	Bacteriology, Cell Anatomy, Membrane Biology, Electrophoresis

Figure 7 – Example Level 4 Categories

Biological phenomena and concept generation

After inspecting the categorized results, 14 promising biological phenomena are selected and listed in Annex A. Discussing these phenomena with graduate-student members of the biomimetic fuel-cell design group of the Biomimetics for Innovation and Design Laboratory at the University of Toronto, two phenomena were chosen for further study:

a) Membrane osmosis (derived from match categories Extremophiles, Membrane Biology) – Membranes are used in cells to enclose a space that differs chemically and biologically from the outside of the membrane. An important feature is the selectively permeable structure that is formed by a lipid bilayer. Substances are allowed to selectively pass through the membrane due to the differences in concentration and properties of the substances (Purves et al., 2001). Figure 8 shows membrane osmosis.

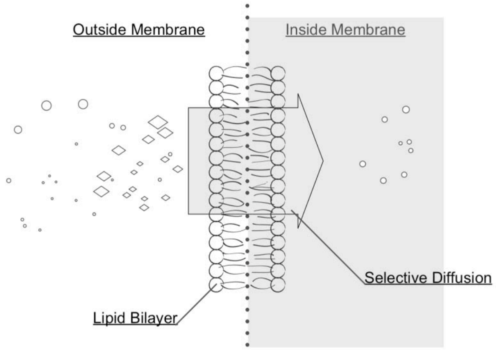

Figure 8 - Membrane Osmosis

b) Water and nutrient transport in plants – Roots, especially the region of maturation, absorb nutrients and water through the soil. Water travels through a system of stems and nodes, finally evaporating from the leaves by transpiration through stoma (Purves et al., 2001), shown in Figure 9.

The proposed redesign based on the above two phenomena places air channels between the fuel/oxidizer flow channels. Air in these channels is kept at a low pressure (~0.5 atm). At this pressure, water evaporates at ~70 ^{o}C, a typical PEMFC operating temperature. A large amount of heat can be removed by evaporating the water generated in the fuel cell reaction. Also, water vapor may be easier to remove than liquid water.

Figure 10 shows the proposed redesign. Water generated at the cathode side is transferred to the low-pressure air channels by osmosis. A selective membrane (e.g., heat-treated porous carbon) prevents oxidizer or fuel from entering the low-pressure air channels. A cross-flow separator prevents fuel/oxide crossing. Figure 11 illustrates how the two biological phenomena were mapped to develop the redesign.

There exist bipolar plate designs related to the proposed concept. An example is an active water management system using an electro-osmotic pump to remove water through a porous carbon bipolar plate. This design has been prototyped with promising results (Litster et al., 2007). Another example is a water transport plate design that uses the pressure difference between the oxidant flow channel and the fuel flow channel to passively move water across the bipolar plate (Yi et al., 2004).

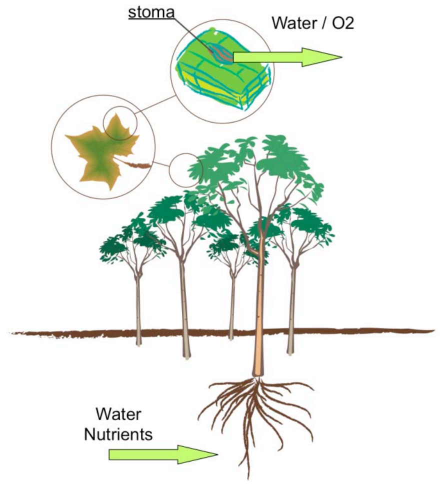

Figure 9 - Water/Nutrient Transport in Plants

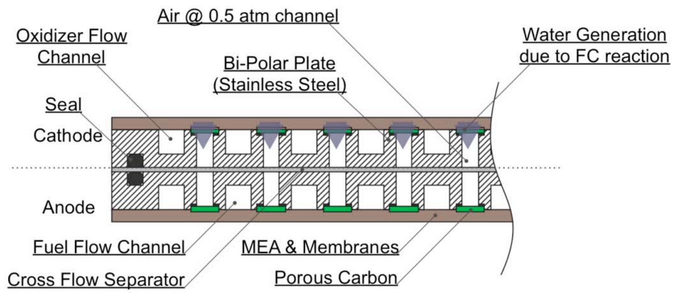

Figure 10 - Proposed Biomimetic PEMFC Bipolar Plate Redesign

Biology Domain		Engineering Domain
Semi-permeable membrane	>	Selective diffusion to air channel
Osmosis enables diffusion	>	Pressure difference between the air channel and oxide/fuel flow channel enables diffusion
Lipid bilayer structure	>	Heat treated porous carbon
Transpire through stomata (Cools & enables mass flow)	>	Evaporate water in the air channel (Cools the fuel cell and enables mass flow of water)

Figure 11 - Mapping from Biological Phenomena to Redesign of Fuel Cell Bipolar Plate

Having related designs, especially those that have been prototyped, provides partial validation for proposed concepts. The key difference between the proposed design and the existing designs is the use of evaporation as the main mechanism for heat and water management. The advantage of this design may include:

1. Evaporation is generally effective in removing heat.
2. A fuel cell system with the proposed bipolar design can be cheaper due to reduced requirement for heat management as the evaporation process removes some heat.

CONCLUDING REMARKS

Arguably, the biological phenomena used in the case study could be identified without the aid of categorization. However, with categorization, the required effort is reduced. The factors that contribute to this reduction are summarized below:

1. For our case study, each match excerpt contains approximately 20 words. There were 405 matched sections, corresponding to 8100 words (20 x 405 excerpts) that may require reading to determine the relevance of each match. With categorization, users may initially choose to rely on the category title to assess the need to view the corresponding matches. As each category title contains about 1.5 words, potentially only 1625 words (1.5 x 1083 categories) are read by using the categorization system, representing up to an 80% reduction over the uncategorized biomimetic search results.
2. Categorization is a process that groups similar ideas. Under this pretext, the results in each category are similar or at least related. Therefore, corresponding biological phenomena can be identified for each category instead for each match. Although this benefit is more difficult to quantify, we observed a reduction in effort when carrying out the case study compared to past case studies using uncategorized search results.
3. The categories may be particularly useful when one is unfamiliar with the subject of search results. One can advance from more general descriptors (level 2 categories) to more specific descriptors (higher numbered categories) of the results. More specific descriptors often correspond to terminology, which once defined, facilitate comprehension of unfamiliar subjects that may be relevant. This benefit is currently limited by noise categories, which will be reduced in future work.

SUMMARY

This paper described the application of a biomimetic search approach augmented with categorization functionality to develop engineering solutions based on biological phenomena. By incorporating a computerized categorization method into the existing biomimetic search tool, users can quickly view summarized results, thus simplifying navigation through a potentially large number of matches. This paper also described how the category structure is extracted from Wikipedia. Search results are displayed in three stages: the category stage, the excerpt stage and the section stage. Each stage increases the depth of the information content but narrows the field of information. Using these display stages as well as increasingly specific categorization levels, users can progressively narrow their search results to identify relevant biological phenomena.

A redesign of a water and heat management system for the PEM fuel cell serves to illustrate this approach. In this case study, the benefits realized of the implemented categorization functionality include reduction in the effort required to identify several relevant biological phenomena from a large group of search results. Seven concepts were generated using the biological phenomena identified, and a concept that combines ideas based on two biological phenomena is currently being further developed to assess feasibility.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support of the Natural Sciences Engineering Research Council of Canada in the form of a Strategic Projects Grant: Biomimetic Redesign of Fuel Cells. 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.

Bio-Medicine – About Us, 2009, http://www.bio-medicine.org/?q=about_index.html (last accessed February 21, 2009).

Bio-Online - Life Science Reference, 2009, http://www.biology-online.org/ (last accessed, May 20, 2009).

Chen, H., S. Dumais, 2000, Bringing Order to the Web: Automatically Categorizing Search Results, *Proc. CHI'00, Human Factors in Computing Systems*, Hague, The Netherlands, pp. 145-152.

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.

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

Gero, J.S. and U. Kannengiesser, 2007, A Function-behavior-structure Ontology of Processes, *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 21:379-391.

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

Hao, T., Z. Lu, S. Wang, T. Zou, S. Gu, L. Wenyin, 2008, Categorizing and Ranking Search Engine's Results by Semantic Similarity, *Proc. 2nd Int. Conf. on Ubiquitous Information Management & Communication*. Suwon, South Korea: ACM New York, NY, USA, pp. 284-288.

Kitamura, Y., N. Washio, Y. Koji, M. Sasajima, S. Takafuji, R. Mizoguchi, 2006, An Ontology-based Annotation Framework for Representing the Functionality of Engineering Devices, *Proc. IDETC/CIE*, Philadephia, PA, DETC2006-99131.

Levitin, D. J., 2002, *Foundations of Cognitive Psychology*. Boston: MIT Press.

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. IDETC/CIE*, Las Vegas, NV, DETC2007-34948.

Litster, S, C.R. Buie, T. Fabian, J.K. Eaton, J.G. Santiago, 2007, Active Water Management for PEM Fuel Cells, *Journal of The Electrochemical Society*, 154:B1049-B1058.

MySQL 5.1 Reference Manual, 2008. http://dev.mysql.com/doc/refman/5.1/en/fulltext-stopwords.html (last accessed May 20, 2009).

Pedersen, T., S.V. Pakhomov, S. Patwardhan, C. G. Chute, 2007, Measures of Semantic Similarity and Relatedness in the Biomedical Domain, *J. Biomedical Informatics*, pp. 288-299.

Porter, M., 2006, The *Porter Stemming Algorithm*, http://tartarus.org/~martin/PorterStemmer/ (last accessed May 20, 2009).

Prescott, L., 2006, *Wikipedia and Academic Research*, http://weblogs.hitwise.com/leeann-prescott/2006/10/wikipedia_and_academic_researc.html (last accessed May 20, 2009).

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.

Schunn, C.D. and A.H. Vera, 1995, Casuality and the Categorization of Object and Events, *Thinking & Reasoning*, 1/3:237-284.

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.

Stone, R.B., and K.L. Wood, 2000, Development of a Functional Basis for Design, *Journal of Mechanical Design*, 122/4:359-369.

Syed, Z.S., T. Finin, A. Joshi, 2008, Wikipedia as an Ontology for Describing Documents, *Proc. 2nd International Conference on Weblogs and Social Media,* Seattle, WA.

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

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

WordNet – Princeton University Cognitive Science Laboratory. 2006. http://wordnet.princeton.edu (last accessed April 26, 2009).

Yi, J.S., D. Yang, C. King, 2004, Water Management along the Flow Channels of PEM Fuel Cells, *AlChE Journal*, 50/10:2594-2603.

ANNEX A – SELECTION OF CATEGORIZED RESULTS FOR THE CASE STUDY

Categories									Section	Section summary
Level 1		**Level 2**		**Level 3**		**Level 4**				
Biology	>	Botany	>	Plant Reproduction	>				29.2.2	Animals may eat these tissues and then disperse the seeds in their feces, often carrying them considerable distances from the parent plant.
Biology	>	Organisms	>	Microorganisms	>	Prokaryotes	>		4.2.2	Enclosing the cell wall and outer membrane in some bacteria is a layer of slime, composed mostly of polysaccharides and referred to as a capsule. The capsules of some bacteria may protect them from attack by white blood cells in the animals they infect. The capsule helps keep the cell from drying out, and sometimes it traps other cells for the bacterium to attack. Many prokaryotes produce no capsule at all, and those that do have capsules can survive even if they lose them, so the capsule is not a structure essential to cell life.
Biology	>	Zoology	>	Animal	>	Poriferans	>		31.5	A sponge feeds by drawing water into itself and filtering out the small organisms and nutrient particles that flow past the walls of its inner cavity. Feeding cells with a collar and a flagellum, called choanocytes, line the inside of the water canals. By beating their flagella, the choanocytes cause water to flow into the animal, either by way of small pores that perforate special epidermal cells (in simple sponges) or through intercellular pores (in complex sponges).
Biology	>	Marine Biology	>	Fish anatomy	>				49.2.1	The fish heart has two chambers. A less muscular chamber, called the atrium, receives blood from the body and pumps it into a more muscular chamber, the ventricle. The ventricle pumps the blood to the gills, where gases are exchanged. Blood leaving the gills collects in a large dorsal artery, the aorta, which distributes blood to smaller arteries and arterioles leading to all the organs and tissues of the body. In the tissues, blood flows through beds of tiny capillaries, collects in venules and veins, and eventually returns to the heart.
Biology	>	Medicine	>	Medical specialty	>	Hematology	>		40.5.5	In the usual (”cold”) fish circulatory system, oxygenated blood from the gills collects in a large dorsal vessel, the aorta, which travels through the center of the fish, distributing blood to all organs and muscles (Figure 40.14a). “Hot” fish have a smaller central dorsal aorta. Most of their oxygenated blood is transported in large vessels just under the skin (Figure 40.14b). Hence the cold blood from the gills is kept close to the surface of the fish. Smaller vessels transporting this cold blood into the muscle mass run parallel to the vessels transporting warm blood from the muscle mass back toward the heart.
Biology	>	Medicine	>	Medical specialties	>	Leukocytes	>		49.4.4	Mammals and birds have lymph nodes along the major lymphatic vessels. Lymph nodes are an important component of the defensive machinery of the body (see Chapter 19)……. The lymph nodes also act as mechanical filters. Particles become trapped there and are digested by the phagocytes that are abundant in the nodes. Lymph nodes swell during infection.
Biology	>	Microbiology	>	Microbial growth and nutrition	>	Extremophiles	>		26.6.1 26.6.3	Halophiles live in very salty environments. Halophiles prevent water loss by increasing internal osmolarity of the cell.

Biology	>	Cell Biology	>				34.2.1	Even in cells with a waterproofed secondary wall, water and dissolved materials can pass from cell to cell by way of structures called pits (Figure 34.8) Pits are interruptions in the secondary wall that leave the thin regions of the primary wall, and thus the plasmodesmata, unobstructed.
Biology	>	Cell Biology	>	Cell anatomy	>	Membrane Biology	>	5.4.1	Diffusion is the process of random movement toward a state of equilibrium. Although the motion of each individual particle is absolutely random, in diffusion the net movement of particles is directional until equilibrium is reached. Diffusion is thus net movement from regions of greater concentration to regions of lesser concentration.
Biology	>	Biological Systems	>	Senses	>	Vision	>	45.4.1	When the 11-cis-retinal absorbs a photon of light energy, its shape changes into a different isomer of retinal-all-trans-retinal. This change puts a strain on the bonds between retinal and opsin, changing the conformation of opsin. This change in conformation signals the detection of light. In vertebrate eyes, the retinal and the opsin eventually separate from each other-a process called bleaching, which causes the molecule to lose its photosensitivity.
Biology	>	Biochemistry	>	Biochemistry methods	>	Electrophoresis	>	11.6.1	After DNA replication has been allowed to proceed for a while in a test tube, the numerous short fragments are denatured from their templates and separated by gel electrophoresis (see Figure17.2).
Biology	>	Ecology	>	Limnology	>	Aquatic Ecology	>	45.2	A colony of corals responds to a small amount of meat extract in the seawater around it by extending bodies and tentacles and searching for food.
Biology	>	Physiology	>	Plant physiology	>			8.5.1 35.1.3	In noncyclic electron flow, light energy is used to oxidize water, forming O2, H+, and electrons. Electrons from water replenish the electrons that chlorophyll molecules lose when they are excited by light.
Biology	>	Physiology	>	Organ systems	>	Respiratory system	>	48.3.1 49.2.1	Many cells lining the airways produce a sticky mucus that captures bits of dirt and microorganisms that are inhaled. This mucus must be continually cleared from the airways. Other cells lining the airways have cilia (see Figure 47.2) whose beating moves the mucus with its trapped debris up toward the pharynx, where it is swallowed.