

Intelligent Search for Biologically Inspired Design

Evangelia
Spiliopoulou

Spencer
Rugaber

Ashok
Goel

Lianghao
Chen

Bryan Wiltgen

Arvind Krishnaa
Jagannathan

Design & Intelligence Laboratory, School of Interactive Computing,
Georgia Institute of Technology

espiliopoulou3
@gatech.edu

spencer@cc.g
atech.edu

goel@cc.ga
tec.edu

chenlh@gat
ech.edu

bryan.wiltgen@
gatech.edu

arvind.krishnaa.j
@gmail.com

ABSTRACT

In Biologically Inspired Design (BID), engineers use biology as a source of ideas for solving engineering problems. However, locating relevant literature is difficult due to vocabulary differences and lack of domain knowledge. IBID is an intelligent search mechanism that uses a functional taxonomy to direct search and a formal modeling notation for annotating relevant search targets.

Author Keywords

Intelligent Search; Biologically Inspired Design; Natural Language Processing.

ACM Classification Keywords

H.5.m. [Information interfaces and presentation]; H.3.3 [Information Search and Retrieval]; D.2.10 [Design].

THE PROBLEM

Engineers are problem solvers, and recently they have been looking more and more at biology as a source of inspiration. Examples include the blades of windmill turbines mimicking the tubercles on flippers of humpback whales, self-cleaning coatings emulating the nanoscale bumps on the surfaces of lotus leaves and water harvesting devices imitating the arrangement of hydrophilic and hydrophobic surfaces on the back of the Namibian beetle.

The idea of deliberately using resources from biology to solve engineering problems is called *Biologically Inspired Design* (BID) or *biomimicry* [1]. From a cognitive perspective, BID entails cross-domain analogies for solving

a target design challenge in engineering by transferring elements of a source analogue from biology.

A major stumbling block with BID is that engineers are novices in biology. Thus, they use the web in order to access biological analogues, which can be found in biology papers. However, this technique is not effective for two reasons. Firstly, engineers and biologists speak different ‘languages’. This makes it very hard for engineers to interpret a biology article or even to form an effective query. The second problem is that existing search engines use keywords to filter their results. One difficulty with this approach is that it will likely not target biology articles. Another is that the results will not be precise—the engineer will have to peruse many suggested papers in order to find relevant candidates for deeper study.

The IBID project is developing an intelligent search solution that aids engineers in locating biology articles addressing a specific engineering problem and in quickly determining whether a retrieved article is in fact relevant to the problem.

To address these difficulties, IBID makes use of faceted search [8] with a controlled, functional vocabulary. It also marks up the resultant papers with tags indicating the specific places in the retrieved article that describe the biological processes.

CONCEPTUAL FOUNDATIONS

Semantic Tagging of Biology Articles

In situ studies of biologically inspired design identified “findability” and “recognizability” of biology articles relevant to a design problem as major challenges: *Findability* pertains to locating biology articles relevant to the design problem, and *recognizability* refers to determining that a given biology article is relevant to the problem. To address these challenges, we have developed a technique for semantic tagging of biology articles based on SBF [4] models (short for Structure–Behavior–Function) of biological systems. Vattam & Goel previously implemented the technique in an interactive system called Biologue [10].

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

Copyright is held by the owner/author(s).

IUI'15 Companion, Mar 29 - Apr 01, 2015, Atlanta, GA, USA

ACM 978-1-4503-3308-5/15/03.

<http://dx.doi.org/10.1145/2732158.2732182>

Controlled experiments with Biologue indicate improvements in both findability and recognizability of biology articles. However, Biologue deals with only a small corpus of biology articles, and the semantic tags are hand-coded. The question becomes to what extent can we identify semantic markers through natural language processing and tag them to help the engineers identify relevant articles on the web?

Functional Modeling with SBF

IBID makes use of SBF models in order to overcome these obstacles in biologically inspired design. SBF is a modeling notation, originally developed for describing mechanical devices. The *Structure* of a system consists of its physical components and substances and the connections among them. The *Behavior* of the system describes the causal mechanism that achieves the system's Function. Finally, the *Function* of a system is an abstraction over externally visible behavior that describes the system's purpose. The idea behind the IBID project is similar to Biologue: SBF models are an efficient way to classify and retrieve biology papers.

Article tagging is enabled by constructing the SBF model of each paper. This tagging is later used during search in order to filter results by any element(s) of the SBF model. This

Some external stimulus (the user plugging in the gun's power cord) causes electricity to flow through the gun. (2) The flow of electricity within the gun causes it to heat up its nozzle. (3) Another external stimulus (the user squeezing the glue gun's trigger) causes the gun to push the solid glue stick against the hot nozzle, which liquefies the glue. (4) Continued pressure causes the gun to extrude the liquid adhesive. Finally, (5) the liquid adhesive comes into contact with the target surface. The Structure of this glue gun includes components, such as the nozzle of the gun, its trigger, its power cord, the chamber of the gun, the glue stick, and the target surface, and substances, such as the electricity that flows through the gun. Structure also describes how the components are connected; for example, that the glue stick is inside the chamber of the gun.

Functional Taxonomy and Natural Language Processing

In order to facilitate the search by engineers of biology articles, we created a common, functional, controlled vocabulary. The key aspect is *function* as both biological organisms and engineering devices perform functions, such as transporting water.

We have developed such a taxonomy of functions that provides a controlled vocabulary common to biologists and engineers. It derives from two existing sources: the Functional Basis [2] and the Biomimicry Institute [3].

Our functional taxonomy is used for two purposes. First, when an engineer wishes to search for a biological mechanism, he or she may use IBID's faceted search interface, which makes use of items selected from the taxonomy to filter potential responses.

The second use of the functional taxonomy occurs when biology papers are analyzed. Of course, these papers use words that are not in the taxonomy. We have therefore constructed an algorithm that makes use of existing Natural Language Processing (NLP) resources, specifically the Stanford Parser, WordNet, VerbNet and PropBank. In brief, the Stanford Parser [5] produces a syntax graph (dependency graph) expressing each sentence of a paper in terms of its parts of speech and their relationships. The parser also determines the *root verb* of the sentence. Our algorithm tries to locate this verb by first looking in the taxonomy and then in VerbNet [9] and WordNet [6]. If the verb occurs only in WordNet, the algorithm tries to find the closest VerbNet verb to it, where *closeness* is determined using synonym and hyponymy (isA) relationships.

The reason IBID's algorithm targets VerbNet is that VerbNet provides a set of expected lexical participants, expressed as frames, for sentential occurrence of each verb. For example, if the verb is *transport*, VerbNet indicates that there should be an agent for doing the transporting and an object being transported. VerbNet is closely associated with PropBank [7], and PropBank provides a set of rules for mapping syntax graphs to VerbNet frames. For example, PropBank can map both active and passive sentences about

IBID - Intelligent Biologically Inspired Design

The screenshot shows the IBID search interface. On the left, under 'Please choose one or more filters: You have chosen:', there are several filter buttons: 'ecology(34)' (green), 'produce(45)' (orange), 'move(80)' (orange), 'classify(23)' (orange), 'marina(2)' (green), and 'environment(13)' (green). On the right, under 'Search Results:', there is a table for 'Water_Transport.txt'. The table has a header 'Open Semantic Analysis Steps' with a plus/minus icon. The table rows are: 'Sentence' (with a 'Show' button), 'Predicate' (with a 'lend' button), 'Thematic Roles' (with sub-rows for Theme: further support, Agent: The results, Recipient: the conclusion), and 'Sentence' (with a 'Show' button).

Figure 1: Faceted search for biology articles.

leads us to IBID's first research hypothesis; SBF elements as indices can improve precision of retrieval versus traditional keyword search.

The second area where SBF tagging is used concerns relevancy checking. It is the second research hypothesis of IBID that SBF models constitute an efficient way to represent an article's content and mechanisms. In this way, SBF tagging can improve an engineer's ability to perform relevancy checks on identified articles.

To better understand functional modeling with SBF, consider a hypothetical electric glue gun. The Function of this glue gun is to apply adhesive to a surface. The glue gun achieves this function with the following Behavior. (1)

IBID - Intelligent Biologically Inspired Design

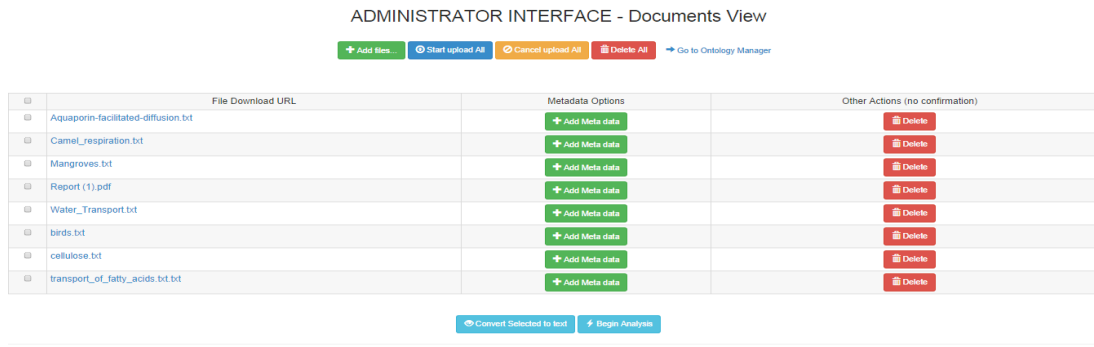


Figure 2. IBID Administrator interface for uploading and analyzing biology papers.

water transport into the same VerbNet frame. Thus, the output of the algorithm is a frame capturing a sentence's content in terms of slot fillers in the frame, thus providing a semantic analysis of the frame's Function.

Here is an example of how the sentence “The pipe transports water from the container” is processed. The algorithm takes *transport* and finds the relevant frames from VerbNet. Two candidate frames correspond to the parses “NP V NP PP.initial_location” and “NP V NP PP.destination”. From these two frames, IBID's algorithm chooses the first as the second cannot map all the words.

Output Frame:

Theme: water

Source: container

Predicate: transport

Agent: pipe

IBID Functionality

IBID supports three use cases: First, IBID enables engineers to search for biology papers that can be used for biologically inspired design solutions to engineering problems. Engineers are able to perform either faceted or keyword search on the engineering problems and receive a sequence of biology papers, sorted by relevance, as results. Figure 1 is a screen capture of the faceted search interface.

In the second use case, administrators can add biology papers to IBID's database, either by directly uploading them or by providing a URL. Administrators are able to analyze any paper in the database and produce an SBF model as a result. Currently, PDF files and text files are accepted, with the PDF files being converted to text for analysis. Figure 2 is a screen capture of an example of what an administrator may see in this use case.

The third use case is when a knowledge engineer wishes to update IBID's functional taxonomy. This situation occurs

when a biology paper contains a verb that cannot be identified using the algorithm described above. Allowed edits include additions, deletions and reorganizations.

IBID IMPLEMENTATION

IBID is accessed by a web browser. On the client side, IBID uses dynamically generated HTML and Javascript. Interaction with the IBID server is via PHP communicating with a MySQL database. Analysis is performed by a Java application that interfaces with the Stanford Parser.

The core element of the application is the Analyzer. It takes as input a biology article in text form and outputs an SBF model of any biological mechanisms found in it. In addition, the database is updated, and various reports are generated. When a user searches for a specific Function, such as *transport*, IBID queries and returns the papers that match with that Function most closely. Selected facets can be used to refine the search to, for example, locate biological mechanisms that transport *water*.

The IBID user interface is generated from information in the database. For example, the Function facet of the search page is populated from the taxonomy stored in the database. Standard AJAX techniques are used to accomplish this.

STATUS

IBID is a working prototype. As we mentioned before, the ultimate goal is to be able to construct a full SBF model from any given text with biology content. These SBF models will then be used for tagging and searching in the database of analyzed papers. Currently, IBID focuses mainly on Functions, which are the crucial component for cross-domain search. We are able to locate and analyze all Functions in any text no matter if it is related to biology. In order to realize this we make use of the resources described in the *Functional Taxonomy and Natural Language Processing* section above. The error rate on the analysis of Functions is relatively low for terms immediately closely related to the verb (like subject/verb relationships). Although it increases in more complicated sentences, we

believe it does not affect the construction of SBF models, for various reasons discussed below.

The main issue with our current result is that we locate and analyze all verbs of every sentence in a text. However, the Function component of an SBF model does not contain all verbs, but only those that determine the content of the text and are part of an underlying Behavior. Consequently, we need a way in order to filter the most important Functions out of all those that the algorithm currently returns. At this point, IBID puts in practice a key-idea of SBF modeling: interaction and dependence between components. For example, we know that VerbNet frames for function verbs refer to objects. As a result, we retain only Functions that refer to relevant structural objects.

Currently we are working on how to determine a relevant set of Structures at the correct level of abstraction. Our idea is to use a frequency heuristic on terms produced from the analysis of Functions. We believe that Structural terms are used many times in a text, something that is in accord with our human experience. Thus, we hypothesize that (a) terms that are used many times in a text probably are Structures and, because of this, (b) our frequency heuristic will generate a good set of candidate Structures. Once we have determined the relevant Structures, we can use them to populate a structural dimension of the faceted search.

EVALUATION

Although the IBID project has not yet completed a formal evaluation, the individual pieces have been tested. Because the Function part is essential, we manually located the Functions of several articles and semantically analyzed them in order to compare the results with our algorithm's output. Our test size contains around 200 Functions from papers with biology content.

The formal evaluation of IBID project as a whole will comprise the following three parts.

- Precision of search: What is the precision of the search results produced by the function-based faceted search as compared with that provided by the keyword search?
- Quality of SBF models: IBID produces SBF models. How well do those models compare with those produced manually by experts? The primary metrics will be completeness and accuracy.
- Interface Effectiveness: How easily can engineers make use of IBID's user interface?

SUMMARY

In this paper, we have presented our preliminary work on an intelligent search engine called IBID that will aid engineers who are conducting biologically inspired design.

What does it mean for search to be *intelligent*? For IBID, intelligence comes from two sources. First is that the

engineer may use domain-aware facets for refining the search query, thereby improving precision across the retrieved papers. Second is that retrieved papers are annotated with tags indicating relevant material. The tags denote textual areas indicative of SBF model elements, such as Behavior mechanisms responsible for providing designated functions. Thus, an engineer can use IBID to perform an intelligent search of biology articles containing descriptions of mechanisms potentially capable of addressing engineering design challenges.

REFERENCES

1. Benyus, J. *Biomimicry: Innovation Inspired by Nature*. William Morrow, (1997).
2. Cheong, H., Chiu, I., Shu, L.H., Stone, R. B. and McAdams, D.A. "Biologically Meaningful Keywords for Functional Terms of the Functional Basis." *Journal of Mechanical Design*, 133(2): 021007-1-021007-11, (2011).
3. Deldin, J.-M. and Megan Schuknecht. "The AskNature Database: Enabling Solutions in Biomimetic Design." In *Biologically Inspired Design*, Chapter 2, pp. 17-27, A. K. Goel et al., editors. Springer-Verlag, (2014).
4. Goel, A.K., Rugaber, S., and Vattam, S. "Structure, Behavior, and Function of Complex Systems: The Structure, Behavior, and Function Modeling Language." *International Journal of Artificial Intelligence in Engineering Design, Analysis and Manufacturing*, 23(Special Issue 1)23-35, (2009).
5. Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S.J. and McClosky, D. "The Stanford CoreNLP Natural Language Processing Toolkit". In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, (2014)
6. Miller, G.A. "WordNet: A Lexical Database for English". *Communications of the ACM*, 38(11): 39-41, (1995).
7. Palmer, M., Dan Gildea, D., and Kingsbury, P. "The Proposition Bank: A Corpus Annotated with Semantic Roles". *Computational Linguistics Journal*, 31:1, (2005).
8. Prieto-Díaz, R.. "Implementing Faceted Classification for Software Reuse." *Communications of the ACM*, 34(5):89-97, (1991).
9. Schuler, K.K. *VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon*. PhD Dissertation, University of Pennsylvania, (2005).
10. Vattam, S. and Goel, A. "Biological Solutions for Engineering Problems: A Study in Cross-Domain Textual Case-Based Reasoning". In *Procs. 21st International Conference on Case-Based Reasoning*, Saratoga Springs, New York, pp. 343-357, (2013).