INTERNATIONAL DESIGN CONFERENCE - DESIGN 2012
Dubrovnik - Croatia, May 21 - 24, 2012.

AUTOMATED CLASSIFICATION INTO THE BIOMIMICRY TAXONOMY

D. Vandevenne, P.-A. Verhaegen, S. Dewulf and J. R. Duflou

Keywords: biomimicry, biologically-inspired design, biomimicry taxonomy

1. Introduction

Biologically-Inspired Design (BID), also known as Biomimicry, Biomimetics, Bio-Inspiration and Bionics, is the discipline where inspiration is taken from nature to solve problems humans encounter. BID is receiving increasingly more attention from research and industry because of the two main advantages the field is often associated with: sustainability and proven performance [Bar-Cohen 2006], [Benyus 1997]. Other noteworthy advantages of biomimetic products are their enhanced marketability caused by the green image of such products, by the savings they often imply (e.g. less energy consumption) and by the association to the organism itself (e.g. to swim as fast as a shark with a biomimetic swimsuit derived from shark skin analysis). Furthermore, drawing inspiration from a largely unused biological knowledge domain entails a higher probability of identifying leapfrog innovations.

These high expectations of biomimetic products are currently not met with adequate methods and algorithms to enable designers to systematically identify candidate biological strategies for biomimetic design. Most existing biomimetic ideas currently originate from spontaneous inspiration, e.g. the invention of Velcro. The inventor, George de Mestral, spontaneously observed the ability of the cockleburs to attach itself to the fur of his dog. This inspired him to study the phenomenon in detail and to develop the well-known innovation. Another way to integrate bio-inspiration into the innovation process is the employment of a multidisciplinary design team, an approach which, although expensive, provides no guarantee for success.

This paper describes the first steps towards a scalable, knowledge-based, systematic BID process based on the Ask Nature [Ask Nature 2011] classification scheme named the Biomimicry Taxonomy. It is an openly available tool that offers support for bio-inspiration during the early stage of the BID process. Although its knowledge base is relatively small and slowly growing, informal experiments with mechanical engineering master students, have demonstrated the usefulness of the original Ask Nature tool. Therefore, automated classification of large numbers of natural language documents containing biological strategy information is proposed to scale the methodology. The paper is structured as follows: Section 2 describes an overview and analysis of the classification scheme used by Ask Nature, Section 3 gives insight into the properties of the Ask Nature knowledge base, Section 4 details an algorithm for automated classification into the Biomimicry Taxonomy, Section 5 illustrates the feasibility of such a classification task and the Section 6 summarizes the contribution into conclusions.

2. Analysis of the biomimicry taxonomy

The Biomimicry Taxonomy is the functional, three-level hierarchical classification mechanism used to organize AskNature's knowledge base in order to make it accessible to engineers, designers, architects, etc. The first level is the group level which consists out of eight categories, e.g. *move or stay put*. Each group level category is divided into subcategories named sub-groups. The two sub-groups for the *move or stay put* group are, for example, *attach* and *move*. Sub-groups are further divided into functions. For example, *temporarily* and *permanently* are two function categories for the sub-group *attach*. In this way, the biological strategy of the octopus using suckers to attach itself can be classified across the three levels as: *move or stay put, attach, temporarily*. It is then up to the person looking for inspiration from nature to formulate his problem into the correct Biomimicry Taxonomy class. To facilitate problem formulation, it is suggested by Ask Nature to focus on verbs to express the problem.

A taxonomy, in general, is an information structure used to classify instances. Hence the Biomimicry Taxonomy is used to classify the biological strategies organisms use to tackle the problems they encounter. In a good taxonomy, instances can be positioned into mutually exclusive, unambiguous categories. However, the classification scheme adopted for the Biomimicy Taxonomy often allows a single strategy to be classified into multiple categories. Take, for example, the harlequin beetle: an organism that uses its strong, large mandibles to escape from the trees in which they are born by chewing through wood. This strategy is currently classified in the following two categories: *break down, physically break down, biotic materials* and *move or stay put, move, in solids.* Although these two example classifications already illustrate the ambiguity of the Biomimicry Taxonomy, the authors propose a third classification: *modify, modify physical state, size/shape/mass/volume.* The first two classifications are more related to the immediate goals for the organism itself, while the last classification is more generally applicable by not specifying the material type as biotic or by not limiting its function to moving through a solid medium. When looking for knowledge transfer from nature, all three classifications can be considered.

When the title of a biological strategy document is carefully chosen to facilitate classification into the Biomimicry Taxonomy, the classification scheme almost seems unambiguous. Take, for example, the strategy s*ilk assembled on demand* which is classified as *make*, *chemically assemble*, *on demand*. However, when reading the short strategy description, more classifications can be identified. One could focus on the attachment on a molecular scale between long protein chains or on the organism scale between the web and a pray; one could position the material properties of the produced silk in a category reflecting the prevention of structural failure or the management of structural forces; one could zoom in at the organ scale and classify to highlight the special storage configuration of the protein chains in the silk glands or how they are being expelled through the spinning duct. Furthermore, other related strategy descriptions focus on the organism level and on the possible functions of the web itself, e.g. capturing animals and sensing signals. Therefore, when trying to classify biological strategy documents (see Section 4), it is important to keep in mind that biological strategies are realized, and thus explained, through the cooperation of functions across the different biological scales (molecules, cells, tissues, organs, organ systems, organisms, etc.), and hence multiple classifications are likely to be possible.

Although one can argue about the word choice of *taxonomy* in the Biomimicry Taxonomy, the fact that biological strategies can be positioned into more than one category does not impede the functioning of the bio-inspiration tool. In contradiction, this property allows a single biological strategy document to be found for different problem formulations or functions. However, it does have implications for automated classification, as explained in Section 5.

3. Analysis of the Ask Nature corpus

In order to devise an automated classification algorithm, as described in Section 4, a reference corpus is required containing biological strategies with classifications. The Ask Nature strategies were downloaded by using the *expand all* functionality of the Biomimicry Taxonomy [Ask Nature 2011]. This resulted in a corpus of 1247 documents. The obtained strategy texts are preprocessed to eliminate common sections present on all strategy pages, resulting in an average word length of 324 words per

document. This can be compared to, for instance, the average number of words used in a set of 160 biological strategy papers from the Journal of Experimental Biology [JEB 2011]: 5382 words. The compactness of the Ask Nature strategy descriptions is likely to be intentional, to maximise their effectiveness as stimuli during ideation. However, when automatically positioning new biological strategy documents into the Biomimicry Taxonomy, larger reference documents would result in a more representative set of document features.
Next, the strategies in the obtained Ask Nature corpus are classified according to the Biomimicry Taxonomy classification on the Ask Nature strategy pages. The obtained top-level category distribution is shown in Figure 1. The figure shows an uneven distribution reflecting that some categories are more likely to be used than others. This bias can be caused by the community or researchers respectively creating the Ask Nature strategies or describing biological strategies, by the nature of the selected classification scheme where some classifications are more evident to consider than others, by the real strategy distribution present in nature per category, or a combination of such factors. Such an uneven category distribution in the corpus, is an important property when considering automated classification.

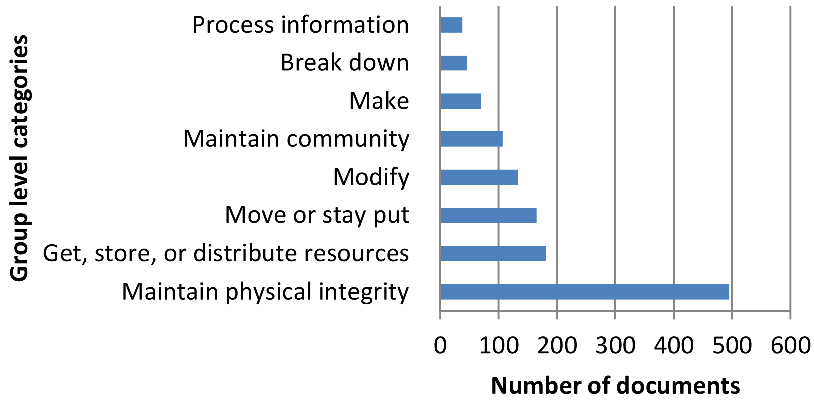

Figure 1. Group level strategy distribution in the acquired corpus

At the deepest level of the three-layered Biomimicry Taxonomy, it should be noted that quite a number of function level classes have very few or even zero strategies assigned to them. Ask Nature explains the latter in their frequently asked questions by their current emphasis on particular sustainability challenges and confirms that its goal is to populate all categories. The knowledge that not all categories have strategies assigned to them has consequences for the development of an automated classifier based on the acquired corpus, as explained in the next section.

4. Classification

The database behind the Ask Nature tool is manually expanded, hence its slow growth. At the time of writing this paper, 1465 strategies are reported and the database has grown with a rate of about one hundred strategies per year during the last two years. Although compared to the existing state-of-the-art in systematic knowledge transfer with nature as source domain [Vandevenne et al. 2011] this number is quite high, it contrasts against the 1.7 million species that are currently named, and against the total number of species that is expected to be between 5 and 30 million [Purves et al. 2001]. To deal with the immense task of manually obtaining and classifying a representative strategy database that reflects nature's potential, Ask Nature calls on the community to expand the database. However, with the current database growth rate, it is unlikely that such a goal can be reached in the near future. In order to scale the approach, the authors propose to automatically classify biological strategy documents into Ask Nature's Biomimicry Taxonomy.
Summary of the main characteristics of the available corpus and the classification scheme to be automated:

- The large number of categories in the Biomimicry Taxonomy (159)

- The categories are not mutually exclusive making the classification task fuzzy
- Ask Nature strategy documents are short, limiting the amount of interesting features to select per strategy
- Strategies in the reference corpus are not evenly distributed across the target categories
- Some categories have zero strategies assigned to them

In order to prove the feasibility of the above described classification task, both the documents of the reference corpus (Ask Nature) and of the sample corpus (see Section 5) are represented according to the Vector Space Model and the cosine distance [Beaza-Yates and Ribeiro-Neto 1999] between the sample and reference documents forms the basis for categorization. The steps of the algorithm are detailed in Figure 2 below.

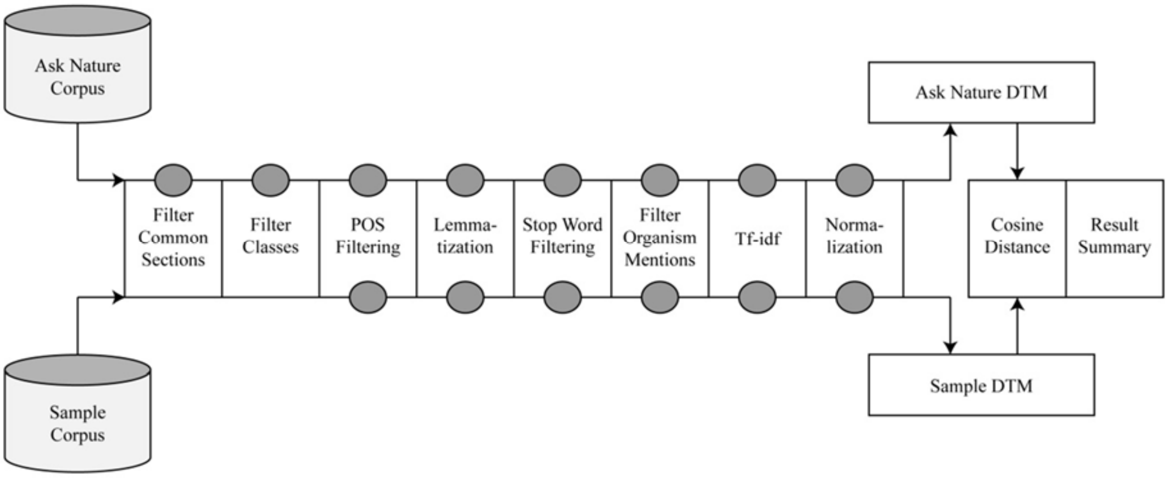

Figure 2. The proposed automated classification algorithm; the dots express which preprocessing steps are performed on which corpus

The **preprocessing** steps that build the Term Document Matrixes which express the Ask Nature and sample corpus in the Vector Space Model are explained below:

- The common document sections for all Ask Nature strategies, such as contact information, are filtered.
- The classifications in the Biomimicry Taxonomy are filtered from the documents to avoid bias to the terms used in the classification scheme.
- Part-Of-Speech (POS) tagging is performed and the verbs, adverbs, adjectives and nouns are retained for further processing.
- The POS information is used to guide WordNet-based lemmatization for the remaining corpus terms. This eliminates all terms that are not inflections of WordNet lemmas.
- Stop words are removed.
- The mentions of organisms are removed to avoid that the classifications are based on organism names instead of terms related to the described biological strategy. Mention detection is performed by an open source species name identification system called LINNAEUS [Gerner et al. 2010]. The database is expanded to include all scientific and common organism names of the NCBI taxonomy, spread over 26 biological ranks, as biological strategies often contain mentions of ranks higher than the species level.
- The corpus is term frequency inverse document frequency (tf-idf) weighted. This gives more importance to terms occurring more frequently in a document and less importance to terms that occur in many documents of the corpus.
- The corpus is normalized to account for the differences in document size.

In order to classify the documents from the sample corpus, the cosine distance [Beaza-Yates and Ribeiro-Neto 1999] is calculated between the documents of both corpora. The **results are summarized** by taking, for each sample document, the Biomimicry Taxonomy classifications of the twenty Ask Nature documents that have the smallest cosine distance to the sample document. These classifications are grouped on the deepest Biomimicry Taxonomy level, the function level, to obtain a candidate classification. In this way, each retrieved classification receives a score on twenty. The results are illustrated in the following section.

5. Results

The ten strategies in Table 1 were randomly drawn from old volumes of the Journal of Experimental Biology [JEB 2011] as these publications were available and likely to contain biological strategies. It was observed that the sample corpus exhibited a bias towards *movement in liquids*, therefore, some documents from this category were discarded in order to obtain a more uniform distribution. As explained in Section 2, for most strategies more than one correct classification in the Biomimicry Taxonomy is possible. Table 1 shows the classification results for the sample corpus. The scores reflect how many of the twenty closest reference documents have the same classification, which is shown in the previous three columns, on all three levels of the Biomimicry Taxonomy. In order to decide whether the candidate classifications proposed by the algorithm are correct, the opinion of the authors was combined. From the ten sample strategies, seven instances were classified into one of the correct categories, three were not as can be seen in Table 1. While making these classifications, the classification algorithm has overcome the strong bias in the Ask Nature corpus towards the group category *maintain physical integrity* (see Figure 1).

Table 1. Classification results for the sample corpus

Strategy document title	Group	Sub-group	Function	Score	Correct
Studies in animal locomotion: the movement of fish with special reference to the eel	Move or stay put	Move	In/on liquids	10	Yes
The lift produced by the heterocercal tails of selachii	Move or stay put	Move	In/on liquids	6	Yes
Medium and long-term changed in the behaviour of visual neurons in the tritocerebrium of locusts	Modify	Modify physical state	Size/shape /mass/volume	3	No
Power requirements for horizontal flight in the pigeon	Move or stay put	Move	In gasses	8	Yes
Bifunctional muscles in the thorax of grasshoppers	Make	Generate/ convert energy	Mechanical energy	6	Yes
The mechanism of orientation of roach, Rutilus rutilus L. in an odour gradient	Maintain physical integrity	Regulate physiological processes	Homeostasis	3	No
The jump of the flea: a study of the energetics	Get, store, distribute resources	Store	Energy	7	Yes
Motor patterns during flight and warm-up in Lepidoptera	Move or stay put	Move	In gasses	7	Yes
Respiratory exchange and evaporative water loss in the flying budgerigar	Get, store, distribute resources	Distribute	Gasses	6	Yes
Pseudo-rheotropism in fishes	Move or stay put	Move	On solids	4	No

In order to demonstrate which terms relate the sample and reference corpus documents, Table 2 presents the most frequent words linking the first strategy in Table 1 with its closest match in the acquired Ask Nature corpus.

If Table 1 is sorted according to the obtained scores, the seven correct classifications separate from the three incorrect classifications. This indicates that the calculated score can be seen as a measure of trust for the candidate classification and that high precision can be obtained through the use of the score as a cut-off value. Such boosted precision would be at the cost of recall, which is of less importance as large amounts of biological strategy documents can be automatically collected with a focused webcrawler [Vandevenne et al. 2012] and the classification method is fully automated.

Table 2. Most important terms linking a sample to a reference strategy document

Verbs	move, swim, travel, generate, increase
Nouns	movement, wave, direction, fin, jet, tail, flow, head, body, force, propulsion, path, motion
Adjectives	efficient, lateral, middle, single, opposite, high, mechanical, large
Adverbs	backward, forward, back, away

6. Conclusions

Ask Nature is a publicly available bio-inspiration tool that aims at providing support during the early design stage. Although the knowledge base of the original Ask Nature tool is relatively small compared to the potential nature has to offer, informal experiments have demonstrated that the original Ask Nature tool does help to generate ideas for BID. In this paper the authors have demonstrated the feasibility of automated classification of large numbers of biological strategy documents into the Biomimicry Taxonomy. The presented approach aims at scaling the bio-inspiration approach of Ask Nature to integrate the large amounts of existing biological strategies much faster than currently is the case. Although the reported experiments are not extensive enough to calculate detailed precision scores, the first results indicate the feasibility of the classification task. In this way, it is shown that it becomes feasible to generate biologically-inspired design ideas from a representative database that reflects the current available human knowledge of nature.

References

Ask Nature, Browse the biomimicry taxonomy, Ask Nature, The Biomimicry Institute, http://www.asknature.org/browse, last accessed on 10/01/2011

Bar-Cohen, Y., Biomimetics: biologically inspired Technologies, 2006 (CRC/Taylor & Francis). ISBN: 0849331633

Beaza-Yates, R., Ribeiro-Neto, B., Modern Informatin Retrieval, 1999 (Addison-Wesley Longman Publishing Co.) ISBN:020139829X

Benyus, J. M., Biomimicry: Innovation Inspired by Nature, 1997 (Harper Perennial). ISBN:0060533226

Gerner, M., Nenadic G. and Bergman, C. LINNAEUS: A species name identification system for biomedical literature. BMC Bioinformatics. 11, 1 (2010), 85

JEB, The Journal of Experimental Biology. http://jeb.biologists.org/, online ISSN: 1477-914. Last accessed at16/12/2011

Purves, W. K., Sadava, D., Orians, G. H., Heller, H. C., Life:The Science of Biology, 2001 (W.H.Freeman & Co Ltd) ISBN: 0716726297

Vandevenne, D., Caicedo, J., Verhaegen, P.-A., Dewulf, S., Duflou J.R. Webcrawling for a biological strategy corpus to support Biologically-Inspired Design. 22nd CIRP Design Conference, 28-30 March, 2012, Bangalore, India.

Vandevenne, D., Verhaegen, P.-A., Dewulf, S., Duflou, J. R., A scalable approach for the integration of large knowledge repositories in the Biologically-Inspired Design process. ICED11, 15-18th of August 2011, Demark

Dennis Vandevenne
Centre for Industrial Management
Department of Mechanical Engineering
K. U. Leuven, Belgium
Email: dennis.vandevenne@cib.kuleuven.be