SEABIRD: Scalable sEarch for systemAtic Biologically-InspiRed Design

D. Vandevenne¹, P.A. Verhaegen¹, S. Dewulf² and J.R. Duflou¹

¹Centre for Industrial Management, Department of Mechanical Engineering, Celestijnenlaan 300A, 3001 Leuven, Belgium

²AULIVE NV Ieper, Belgium

Corresponding Author:

Name: Dennis Vandevenne

Address: Centre for Industrial Management, Department of Mechanical Engineering,
Celestijnenlaan 300A, 3001 Leuven, Belgium

Telephone number: +32 16 37 26 91

Email address: dennis.vandevenne@kuleuven.be

Content Details

Short title: *SEABIRD*

Number of pages in the manuscript: 54

Number of figures in the manuscript: 7

Number of tables in the manuscript: 6

Title

SEABIRD: Scalable sEarch for systemAtic Biologically-InspiRed Design

Abstract

As more and more people are increasingly turning to nature for design inspiration, tools and methodologies are developed to support the systematic bio-ideation process. State-of-the-art approaches struggle with expanding their knowledge bases because of interactive work required by humans per biological strategy. As an answer to this persistent challenge, SEABIRD, Scalable sEarch for systemAtic Biologically-InspiRed Design, is proposed. SEABIRD is based on two conceptual representations, Product and Organism Aspects, extracted from respectively a patent and a biological database, that enable leveraging the ever growing body of natural-language biological texts in the systematic bio-inspired design process by eliminating interactive work by humans during corpus expansion. SEABIRD's search is illustrated and validated with three well-known Biologically-Inspired Design cases.

Keywords

Bio-inspired design, Biologically-Inspired Design, Biomimicry, Creativity & ideation

1 Introduction

Biologically-Inspired Design (BID) is the discipline where inspiration is taken from the natural world to solve technical problems or challenges. An increased interest for BID during the last two decades caused a strong rise in the number of academic papers (Lepora et al., 2013) on a wide range of topics, such as robotics, artificial intelligence, multifunctional materials, sensors, etc. In addition, a growth in patented inventions is observed (Bonser, 2006). While the growth in academic output is an indicator of overall research activity and interest in BID, the increase in patent output suggests that many of the resulting bio-inspired solutions are deemed commercially viable and that new technologies are derived from biological examples (Bonser & Vincent, 2007). Today, BID is even becoming an important paradigm for disciplines like robotics and materials science (Lepora et al., 2013).

Many of the challenges posed upon organisms in their natural environments, e.g. making strong materials, moving efficiently through fluids, implementing shock absorbency, regulating temperature, etc., are similar to the problems humans face. One common rationale motivating Biologically-Inspired Design recognizes the relevance and proven performance of biological solutions (Bar-Cohen, 2006, 2011): Why reinventing the wheel when there are millions of different species around us that adopt time-tested solutions? A second important motivation for nature as a source for inspiration is the general increase in environmental consciousness (Bonser & Vincent, 2007), supported by academic research (Bajželj et al, 2013) and slowly influencing governments (IPCC, 2007). Indeed, organisms rely on renewable resources for their “production processes” and nature does not generate ever-growing waste piles. Although bio-inspired products are not inherently sustainable (Vandevenne et al., 2012a), nature is regarded as a promising source of inspiration for environmentally-friendly products and processes (Gebeshuber et al., 2009) and, by some, even regarded as a measure or ecological standard to judge the "rightness" of innovations (Benyus, 1997). A third argument for BID claims that drawing inspiration from a largely unused biological knowledge domain entails a higher probability of identifying leapfrog innovations. In an experimental setup, exposure to biological examples has been found to increase novelty without decreasing variety in idea generation (Wilson et al., 2010). Other

possible advantages of bio-inspired products are their enhanced marketability and financial savings through efficient use of energy and other resources.

These high expectations of Biologically-Inspired Design are currently not met with adequate methods and algorithms that enable designers to systematically leverage nature's potential. Although it is hard to obtain trustworthy information about the specific mechanisms behind the bio-inspiration for many bio-inspired designs, it is commonly accepted that spontaneous, accidental inspiration plays an important role. A frequently used example is Velcro. Its inventor is claimed to have serendipitously observed the ability of the cocklebur to attach to the fur of his dog; which then inspired him to study this phenomenon in detail and to develop the well-known innovation.

In the last decade, a number of research efforts have focused on providing support in the Biologically-Inspired Design process by developing tools and methods that facilitate cross-domain search and knowledge transfer. Many successes and insights are reported (see Section 2), although there is one unresolved challenge: the scalability of these systematic BID ideation tools and methods. Currently, about 1.7 million species are named and the total number is estimated between 5 and 30 million (Purves et al., 2001). Although only a fraction of these 1.7 million identified organisms has been studied in detail, there exist many sources, such as books, journals and online resources, where biological knowledge is documented. Considering the large work that lays ahead for biologists to completely describe and comprehend all of nature's phenomena, these sources are expected to keep on expanding. To leverage this ever-growing source of biological inspiration in natural-language format, the authors of this paper developed a series of methods and algorithms that support systematic and automated identification of biological information relevant to specific design problems. These methods and algorithms are jointly referred to as SEABIRD, Scalable sEarch for systemAtic Biologically-Inspired Design. In Section 2 the state-of-the-art in systematic search for bio-inspiration is listed and discussed. Next, SEABIRD's architecture and functionalities are detailed in respectively Sections 3 and 4. Thereafter, the proposed approach for scalable search is validated in Section 5, discussed in Section 6 and conclusions are drawn in Section 7.

2 Related research

Biologically-Inspired Design has been studied from a number different perspectives, focusing on application domains (Bar-Cohen, 2011; Bhushan, 2009), on sustainability of biomimetic products (Benyus, 1997), on understanding bio-inspired Design-by-Analogy (Mak & Shu, 2008; Helms et al., 2009; Vattam et al., 2010a; Cheong, 2012a) and on developing ideation systems for Systematic BID (SBID) (see Section 2.2). This paper focusses on function based BID to solve technical problems, not on the mimicry of form or structure for e.g. aesthetic purposes. The remainder of this section first summarizes the four general phases of the typical Scalable Systematic BID process (SSBID), and then presents the state-of-the-art in the search phase to illustrate a common challenge.

2.1 Scalable Systematic BID process phases

A comparison of different contributions describing the encountered phases of the Biologically-Inspired Design process, identifies four ubiquitous phases (Sartori, 2010): (1) formulating search objectives, (2) searching for biological analogues, (3) analyzing biological analogues and (4) knowledge transfer. These four phases are applied to the Scalable Systematic BID process that draws inspiration from large biological repositories in natural-language format (Vandevenne et al., 2011):

- **Formulate search objectives.** The specific design problem at hand needs to be captured in a form allowing launching search in the next phase.
- **Scalable search.** Searching large repositories of biological strategy documents requires a scalable, automated approach that allows the identification of a number of candidate biological strategy documents to be considered in the next phase.
- **Filter and analyze.** An automated search method unleashed on large repositories is expected to generate more than a few candidate descriptions of biological strategies. Therefore, methods and algorithms are required to guide the designer in selecting one or more candidate solutions. The retained biological candidates should be analyzed in detail to enable the designer to transfer the biological principles in the next phase.

- **Knowledge transfer.** In order to transfer knowledge from the biological source domain to the technological target domain in a systematic way, assistance to the designer is required for identifying the cross-domain analogy and to come up with a feasible, biologically-inspired technical concept.

The above Scalable Systematic BID process is expected to be iterative. For example, after attempting knowledge transfer based on a first set of search results, users might change their problem formulation to perform search again. To support one or more of the four general BID process phases, a number of systematic BID approaches have been proposed. The next section summarizes these approaches, while focusing on scalability of the search phase as this is where SEABIRD's main contribution currently lies. A more detailed description of these approaches, positioning each in the above four SSBD process phases, is available in Vandevenne, 2014c.

2.2 Searching for biological strategies

As a summary of the state-of-the-art in the SBID search phase, this section first details two keyword-based search methods, next two approaches supporting on the classification of biological strategies and finally three contributions that require complex model instantiation for each corpus entry.

Starting from a functional keyword search, an iterative and interactive methodology extracts new biological keywords from the obtained results for future searches (Lenau et al., 2010). In this way, biological search words, initially not known to be relevant to the problem, are identified. A contribution aimed at automating the identification of biologically relevant search words (Chiu & Shu, 2007; Shu, 2010) bridges the terminology gap between the engineering and biological domain by means of a systematic, semi-automatic search method that requires the design problem to be expressed in functional keywords; and then generates biological meaningful bridge verbs and text passages containing them. Another WordNet-based contribution—named BIOscrabble—performs search on and with PubMed, a very large biomedical research article database, by entering keyword combinations and inferred keywords (synonyms, variations and negations) in PubMed's search

engine which is used as a block box (Kaiser et al., 2012). Besides function, the approach also formally investigates property and environment related search words (Kaiser et al., 2014).

Two contributions require positioning each biological strategy into a classification scheme. First, BioTRIZ (Vincent et al., 2006), aims at integrating biological knowledge in the TRIZ methodology (Altshuller, 1984) by interactively positioning biological strategies in the BioTRIZ contradiction matrix. To identify bio-inspiration, the problem needs to be formulated as a classical TRIZ contradiction, which is then reformulated into a BioTRIZ contradiction. This BioTRIZ contradiction then leads the designer to inventive principles learned from the manual analysis of 2500 contradictions in 500 biological phenomena. Second, AskNature interactively classifies biological strategies in a functional, hierarchical taxonomy called the Biomimicry Taxonomy (Deldin & Schuknecht, 2014). Designers looking for bio-inspiration need to formulate their design problem in this taxonomy. Recently, a classification algorithm is proposed that automatically positions biological strategies in the functional categories of AskNature's Biomimicry Taxonomy (Vandevenne et al., 2014a).

Three approaches can be found in literature that require human interaction to instantiate detailed models for each biological phenomenon to be integrated in a structured knowledge base. To enable search, the technical problem is also expressed by instantiating at least part of these models. Next, matching of the technical problem model to the biological system models in the knowledgebase generates candidate stimuli for design-by-analogy. Such a methodology has currently been reported for Structure-Behaviour-Function (SBF) models (Vattam et al., 2010b; Goel et al., 2012), for Functional Basis (FB) models (Nagel et al., 2010; Nagel & Stone, 2012), and for SAPPhIRE models of causality (Chakrabarti et al., 2005; Sartori et al., 2010). These three model-based approaches have been recently extended in the following ways. Biologue, a social citation cataloguing system, is developed (Vattam & Goel, 2011) to involve more people in the process of manual creation of SBF models. An Engineering to Biology Thesaurus (Cheong et al., 2011; Nagel & Stone, 2012), a lookup table that translates the Functional Basis terms into biological corresponding terms, is proposed to

extend the Functional Basis approach. This way, model instantiation, requiring interactive work by humans, for large biological databases is avoided as the biological corresponding terms can be used as search words in natural-language texts. The SAPPhIRE approach is extended by an ontology aimed at providing extra stimuli during bio-inspired ideation. The ontology consists of manually derived, biological and engineering term clusters for each of the SAPPhIRE model constructs (Srinivasan et al., 2012).

In order to illustrate the corpora integrated into the above systems, for each contribution, the number of reported biological sources is given in Table 1.

<Table 1>

While these systematic BID approaches differ in many ways, they all assign a central role to function: functional problem formulation and function driving search. A second observation is the increasing interest in natural-language resources. Pioneered by researchers from the Biomimetics for Innovation and Design Laboratory (Chiu & Shu, 2005), later the interactive functional keyword search approach (Lenau et al., 2010), the Engineering to Biology Thesaurus keyword search (Nagel & Stone, 2012) and BIOScrabble (Kaiser et al., 2012) were reported. Using natural-language resources avoids the immense interactive work of populating structured databases, i.e. model instantiation or strategy classification. Therefore, the SEABIRD search system described in this article uses natural-language biological and patent texts as corpora to respectively represent biological and technical systems.

2.3 Challenges for scaling search

All of the above methodologies struggle in one way or another with scalably integrating large numbers of biological systems; these challenges are:

- **Interactive result filtering.** For the above natural-language keyword search methods, interactive result relevance filtering does not scale well for large repositories. For example, an analysis—performed by students—of 3416 research articles from the BIOScrabble approach reports that 3.36% of the returned results were considered inspiring (Kaiser et al., 2014). A

recent approach to identify causally related functions could assist here to return proportionally more relevant results (Cheong & Shu, 2014).

- **Interactive classification.** This translates to the positioning of biological strategies into the Biomimicry Taxonomy for the AskNature approach, or to the identification of the relevant contradiction for the BioTRIZ approach. These interactive tasks are again proportional to the size of the biological databases.
- **Interactive model instantiation.** Model-based approaches are inherently difficult to scale as they require a detailed analysis of both the engineering and biological systems to express them on a common abstraction level.
- **Crowdsourcing.** The SBF-based approach developed a social citation cataloguing system for annotating research articles with model instantiations. In theory, crowdsourcing can tackle the scalability of any BID ideation system. However, the successful creation of a large manually annotated database has not yet been reported. AskNature is demonstrating constant growth of about 100 biological strategies a year. Although most of the strategies in their database is added by paid staff (Deldin & Schuknecht, 2014), a number of qualified scientific curators are also able to add content.
- **Completeness of thesaurus or ontology extensions.** Questions rise about the completeness of the relatively short biological word lists; which, in turn, makes it difficult to estimate how much of the biological inspiration in natural-language texts is retrievable.
- **Reference corpus building.** Automated classification of biological strategies into the Biomimicry Taxonomy requires a certain amount of reference strategies to be manually annotated for each functional category. Achieving adequate reference corpus support for each of the 162 functional categories is feasible but still requires a significant interactive effort (Vandevenne et al., 2014a).

3 System Architecture

The proposed scalable search approach leverages the knowledge about technical systems captured in patents and biological systems documented in academic papers. These two databases allow the generation of two domain-specific concept sets, named Product Aspects (PAs) for technical systems and Organism Aspects (OAs) for biological systems. These aspects are generated by the analysis of word co-occurrences in respectively a patent and a biological document set. For example, the word *hovering* co-occurring together with *flying* in one document can be linked to the word *gliding* co-occurring in another document with the common word *flying*. *Flying* could be an Organism Aspect in this simple example, associated with the words *flying*, *hovering* and *gliding*. Large-scale application of this principle to both corpora results in technical and biological concepts; where each concept is associated to a number of co-occurring words. These concept sets, PAs and OAs, are central to the proposed approach as illustrated by Figure 1. They represent a common abstraction level that enables technical systems to be linked to biological systems.

<Figure 1>

The extraction of Product Aspects from a patent database is fully described in (Verhaegen et al. 2009, 2011). PAs support the PAnDA (Product Aspects in Design-by-Analogy) tool that identifies candidate products for Design-by-Analogy. Using PAnDA, a product, e.g. *carburetor*, can be associated to the most relevant Product Aspects, e.g. for the carburetor case *combustion*, *inflow/outflow*, *rotate*, etc. Based on a selected Product Aspect, e.g. *inflow/outflow*, candidate products are depicted as stimuli for the redesign of the carburetor product, e.g. *strainer*, *faucet* and *euphonium* are suggested as candidate products for design-by-analogy. The PAnDA tool has been shown to increase the variety and novelty of ideas (Verhaegen et al., 2011b).

Analogously to the extraction of Product Aspects from a patent database, Organism Aspects are extracted from a biological database. Because OA extraction is inspired by PAs extraction, there is quite some high-level similarity in the involved processes. Nevertheless, it is necessary to detail the different algorithms as there are a number of necessary implementation differences. First, in Section 3.1, both the technical and biological knowledge bases are detailed. Thereafter, in Section 3.2 the preprocessing algorithms are explained that transform the biological corpus into a Document-Term

Matrix (DTM). Next, in Section 3.3 Organism Aspects are generated from this DTM, and in Section 3.4, an algorithm for automated mapping between PAs and OAs is presented that allows linking biological solutions to technical problems. Finally, in Section 3.5, the Product and Organism Aspects are used to characterize products and organisms.

3.1 Corpora

The corpus representing the technical domain consists of 155,000 patents, randomly drawn from the 21 million patents of the EPO Worldwide Patent Statistical Database (PATSTAT). The full-text descriptions of these patents are used for Document-Term Matrix generation (see Section 3.2) because it has been shown that the inclusion of significantly large text fragments of the description can be beneficial for text-mining in a patent environment (Larkey, 1999; Fall et al., 2003). Starting small and doubling the database size each time while comparing the generated Product Aspect sets, has shown that the extracted concept sets change very little once a significant sample corpus is reached (Verhaegen & Duflou, 2013). In the case of PAnDA, 155,000 patents were found to be sufficient to form a stable conceptual representation of the technical domain. After PA generation and system setup, new patents can be integrated by applying folding-in (Deerwester et al., 1990) to update PAnDA. This folding-in procedure allows positioning new document vectors in a previously created, stable PA vector space. The mathematical procedure is the multiplication of the document vector with the term-PA matrix.

The corpus representing the biological domain, for the tests reported in this paper, is a set of 8011 full-text biological papers from the Journal of Experimental Biology. As much as possible of the available papers from this journal are taken as test set, no further selection was made. In this paper the term *strategy* refers to the biological phenomenon with potential for knowledge transfer. The term *strategy document* refers to a single document describing a strategy. Hence it is possible to encounter multiple strategy documents discussing different aspects of the same strategy. Although the current number of strategy documents is not high enough to confidently claim that a representative subset of human's knowledge about nature is gathered, the research presented here found the resulting Organism Aspect

set to be useful and detailed enough to allow testing and validation of the proposed approach. More details about corpus expansion are discussed in Section 6.

3.2 Document-Term Matrix generation

Five preprocessing steps transform each corpus into a Document-Term Matrix (DTM). The DTM is the Vector Space Model (VSM) (Salton et al., 1975) representation of the corpus. A VSM representation was, for instance, also used to develop a patent-based analogy search tool for innovative concept generation (Murphy, 2011). This algebraic model—the VSM—represents documents as vectors, where each dimension corresponds to a unique corpus word or feature; and each feature value corresponds to the importance of the word in the document. A DTM consists of a collection of document vectors, see Table 2 for a simple example of a document vector. An overview of the DTM generation preprocessing steps, which are detailed in the following subsections, is shown in Figure 2.

<Figure 2>

3.2.1 Organism filtering

To avoid that Organism Aspects would represent parts of the Linnaean taxonomy—a particular form of biological classification created by Carl Linnaeus (Linnaeus, 1767)—the occurrences of organism names in the texts are filtered. Organism name filtering is comparable with filtering words related to products (Verhaegen et al., 2011a) as in both approaches inter-document links are removed to bring out structure relevant for design-by-analogy. Omitting organism name filtering could result in, for example, an Organism Aspect that represents the order of Araneae or spiders, caused by the many mentions of different spider species that tend to co-occur in biological papers. Generating such Organism Aspects would be useless for the envisaged mapping of Product Aspects to Organism Aspects.

Organism name detection is performed by LINNAEUS (Gerner et al., 2010), an open source species name identification system. Its database, containing only names at species-level, is expanded to include all scientific and common organism names of the National Center for Biotechnology Information (NCBI) taxonomy. As biological strategy documents often

contain mentions of ranks higher than the species level, all 26 biological ranks are included (Vandevenne et al., 2014b). For the biological corpus of 8011 documents, on average 128 organism mentions are detected per document and 19782 unique organisms are found in the full corpus. Organism mention filtering eliminates a large number of nouns and therefore reduces the dimensions of the DTM and a significant part of the effort of manual filtering (See Section 3.2.5).

3.2.2 Part-Of-Speech tagging

A standard Trigrams'n'Tags (TnT) tagger (Brants, 2000) annotates each biological text with Part-Of-Speech (POS) information (Charniak, 1997) and only verbs, adverbs, adjectives and nouns are retained for further processing. These are the same POS categories as used for the PAnDA tool (Verhaegen & Duflou, 2013). These POS categories contain terms about function, properties and environment, all relevant for search in SBID as motivated by Kaiser et al. 2014. Table 2 illustrates the POS tagging results of the following sentence: "This study investigates the aerodynamic and gravitational forces on ideal falcons and uses a mathematical model to calculate speed and acceleration during diving.". The table lists all words of the sentence that received one of the above mentioned POS categories in column 1 and their grammatical function in the sentence, or POS category, in column 2. After Part-Of-Speech tagging, corpus documents are treated as bags of words, meaning that their content is represented as an unordered set of words, ignoring word order and lexical relationships.

<Table 2>

3.2.3 Lemmatization

After Part-Of-Speech tagging, each remaining word is reduced to its lemma by WordNet-based lemmatization (Fellbaum, 1998) assisted with the words' POS tags. Table 2 illustrates lemmatization for the running example in column 3. For example, the word *investigates* is lemmatized to *investigate*. By eliminating all words that are not inflections of a WordNet

lemma, this step is also a thesaurus filter that ignores wrongly spelled words. Lemmatization further summarizes the document vectors by summing word frequencies of inflections of the same lemma. Hence, the final DTM is reduced. For example: the word *swimming* occurring 8 times in a document can be linked to the word *swims* occurring 4 times in the same document through the association of both words with the lemma *swim*, and the frequencies of the two features *swimming* and *swims* can be summed to 12 for the lemma *swim*. In order to enable the use of standard terminology like document-term matrix, instead of document-lemma matrix, from here on, the concept *term* will refer to lemmas.

3.2.4 Stop word filtering

Stop words (Fox, 1989) are removed in the fourth filtering step, as they do not represent relevant document content. Examples of stop words are *the*, *and*, *a*, *that*, *was*, *etc*. The full list used in the proposed system can be consulted in (Fox, 1989).

3.2.5 Relevance filtering

All previous preprocessing steps have reduced the number of terms, and hence have decreased the size of the DTM. Manual filtering identifies those terms in the remaining corpus dictionary that are interesting for knowledge transfer from the biological source domain to the technical target domain. Term filtering is performed by considering the terms independently, hence no context in the form of sentences or documents is provided. Four term categories are immediately dismissed: a small number of remaining organism mentions, all terms representing academic language (e.g. study, pose, hypothesis, objective), all terms representing structure (e.g. wood, body, larva, droplet, oil, wing, antenna) and all terms that belong to non-relevant topics (typically related to the technical domain, medicine and genetics). From the remaining terms only those that are interesting for describing a biological strategy for knowledge transfer to the technical domain are retained. For the running example,

study, investigate, mathematical, model and *calculate* are examples of words used in academic language that can be dismissed, as illustrated by column 4 in Table 2.

As such manual term relevance filtering is a subjective process, its repeatability needs verification. Therefore, a chance-adjusted measure of agreement is calculated. The free-marginal multirater kappa (Randolph, 2005) is used since coders are not forced to mark a specific proportion of the terms as relevant. The free-marginal bi-rater Kappa is calculated to be 0.73. Kappa values range from -1 to 1: -1 indicating perfect disagreement below chance, 0 indicating agreement equal to chance, and 1 indicating perfect agreement above chance. As a rule of thumb a kappa value of 0.7 or above is considered proof of good agreement (Randolph, 2005).

Only a relatively small number of words in the English language covers most of the written text. Zipf's law (Zipf, 1932) confirms this by stating that few words occur very often while most words occur rarely. This Zipf function is a negative exponential: see Figure 3 for the term frequencies of the remaining corpus terms after stop word filtering. The most frequent term, not filtered by one of the previous preprocessing steps, is *reference*, occurring 7974 times which is nearly once per document. As it can safely be assumed that sporadically occurring terms contribute little to the formation of the principal components for Organism Aspect generation (see Section 3.3), only the terms occurring more than ten times in the corpus are interactively evaluated. This way 49% of the dictionary is eliminated, almost halving the required interactive work. Relevance filtering further reduces the term dictionary, resulting in a DTM with 8011 rows (documents) and 4055 columns (terms). The six remaining words indicated in the last column of Table 2 represent the document vector for the one-sentence example document. For this very short document, the vector only has six attributes (one for each word) and the term frequencies are all one as every word occurs exactly once in the sentence.

<Figure 3>

3.3 Product and Organism Aspect Generation

The OAs generation process is depicted in Figure 4. The Document-Term Matrix obtained from preprocessing is a matrix where each element a_{ij} corresponds to the frequency with which term j occurs in document i . This matrix is weighted with a Term Frequency Inverse Document Frequency (tf-idf) scheme (Salton & Buckley, 1988) and normalized to account for different document text lengths. The tf-idf weighted Document-Term Matrix is subjected to Principle Component Analysis (PCA) (Berry et al., 1995; Skillicorn, 2007). This analysis allows extracting a predefined number of Principle Components (PCs), of which the first principal component is the dimension oriented in such a way that it explains the maximum amount of variance in the data set. Each succeeding principal component represents as much of the remaining variability as possible, taken into account that all principal components are orthogonal to each other. The authors currently calculate 300 principal components (Landauer & Dumais, 1997; Verhaegen & Duflou, 2013), which is an arbitrarily chosen number for proof-of-concept purposes. This dimensionality reduction results in two smaller matrices: a term-PC and a PC-document matrix.

<Figure 4>

In a tf-idf weighted Document-Term Matrix, term frequencies are correlated variables. For example, documents containing a high frequency of the term *eating* are more likely to contain terms like *feeding* or *ingesting* than random documents. In a term-PC matrix all terms are expressed in a smaller number of uncorrelated variables or principal components. Furthermore, for the PAnDA tool it has been demonstrated (Verhaegen et al., 2009, 2011a) that Varimax rotation (Kaiser, 1958) facilitates the interpretability of the resulting principal components. After rotation, the PCs are called Organism Aspects (OAs), which are represented by a number of ranked terms. Figure 5 illustrates the highest scoring terms for an example OA: *buoyancy*, *descent*, *ascent*, *buoyant*, *depth*, *density*, *ascend*, *descend*. Manual labelling interprets such ranked groups of terms; for example, the OA label *buoyancy*

is given to this group of terms. Table 3 details the OA labels for the first 20 Organism Aspects. As such labelling is subjective, the process is repeated with two raters who have 93.33% agreement. All 300 generated OAs are labeled and 30 random OAs were selected for the estimating the interrater reliability. Although the raters could use any labeling they seemed fit, most of the OA labels were given by picking one of the high-loading terms—or a modified form of one of these terms—on the OAs. For this inter-rater reliability test, the number of possible labels is very large, hence there is no need to adjust for chance. OAs generation results in a term-OA and an OA-document matrix that allow the positioning of terms and documents in the new set of labelled Organism Aspect dimensions. These two matrices are important information structures that support the functionalities described in Section 4.

<Table 3>

<Figure 5>

3.4 Product and Organism Aspect mapping

In order to link technical and biological systems for potential knowledge transfer, a similarity measure between the Product Aspects as specified by (Verhaegen et al., 2011a) and the above calculated Organism Aspects is necessary. The multiplication of the term-PA and the term-OA matrices, with the term indices as common dimensions, results in a PA-OA matrix that expresses the similarity between the technical and biological concepts. The more the PA and OA vectors share common term loadings, their higher similarity value will be in the PA-OA matrix. One can distinguish 3 types of relationships between PAs and OAs:

- *(Near)identical concepts:* The (near) identical concepts are represented by the largest values in the PA-OA similarity matrix. Some examples are *rotation, fluorescence, solar and inflation/deflation*. Each of these concepts exist in both the OA and PA set and a high value links them in the similarity matrix. The automated mapping of (near) identical concepts still allows domain specific terminology to be mapped as it is not necessary that all the terms or features in the (near) identical Product and Organism Aspect vectors are identical.
- *Semantically related concepts:* Some links between PAs and OAs express a semantic relation less obvious than for identical concepts. These links can represent a cross domain bridge on

the conceptual level as, for instance, the PA *drilling* that has a strong link to the OA *digging* and the PA *humidify* has a strong link to the OA *transpiration*. Of course the cross domain bridge on the terminology level is also present here, connecting domain-specific terms related to both concepts.

- *Unrelated concepts:* As can be expected, most links between PAs and OAs are meaningless, which results in a zero or near zero value in the PA-OA matrix.

Aspect mapping associates concepts from the technical domain to the biological domain and, hence, patents to biological papers. In the next section, products are linked to patents and organisms to biological papers.

3.5 Characterization of Products and Organisms

The tasks of characterizing products with Product Aspects and organisms with Organism Aspects are similar. First, the occurrences of products in patents and of organisms in papers needs to be identified and expressed in product-patent and organism-paper matrices. Next, the document vectors in these matrices are folded-in (Deerwester et al., 1990) to obtain product and organism characterization matrices.

3.5.1 Product Identification

The product identification used by the PAnDA tool (Verhaegen et al., 2011a), which consists of 1011 single-word products extracted from the Google product taxonomy (Google Merchant Center, 2014), is improved by a new multi-word product identification algorithm, detailed in the paragraphs below. Motivation is two-fold: on the one hand this way much more products are detected (151132), on the other hand products can be characterized more precisely. For example, products like *air bag* and *plastic bag*, that would be mapped to the single-word product *bag*, are distinguished by the new multi-word product identification algorithm. Such more refined product search improves product characterization results. For example, the most important Product Aspects for *air bag* and *plastic bag* are respectively *inflation/deflation* and *conveying and feeding*, while the most important PA for *bag* is *packaging and sealing*. Without multi-word product identification, some of the validation cases

presented in Section 5 would not be possible, i.e. for the *air conditioning system* and *car body* products.

To identify products, the following Part-Of-Speech sequence is detected in the title and abstracts of patents: zero or more adjectives followed by one or more nouns. Next, unique nouns occurring in the title and abstracts are manually categorized into four categories (see Table 4). The first category represents nouns that are products by themselves, e.g. *valve*, *display*, *vehicle*, *etc.* The single-word products from the Google taxonomy are also placed in this category. A second noun category groups words that need explanation to be products, e.g. *system*, *device*, *apparatus*, *etc.* These are not discarded as the language used in patents often is indirect, like for example, *communication apparatus* or *packaging system*. A third category contains nouns that have an explanatory function, e.g. *temperature*, *pressure*, *power*; these can explain the words in the previous category, e.g. *temperature regulation system*. All other nouns are considered not relevant and placed in the fourth noun category. Adjectives are labelled relevant (e.g. optical, magnetic, digital, etc.) or not relevant (e.g. preferred, first, improved, etc.) to reflect their capability of providing useful information about a product. Manual annotation of the 10% most frequent nouns and 20% most frequent adjectives (in total 2568 words) resulted in the identification of 151132 multi-word products, which is satisfactory for proof-of-concept purposes. This way, rarely used nouns and adjectives (e.g. thromboembolism, saussurea, arrhenius, gratuitous, sovereign, etc.) are filtered as the large majority of these infrequently encountered words would be labelled non-relevant anyway. Noun and adjective categorization is a subjective process, hence rater agreement is measured. The free-marginal multirater kappa (Randolph, 2005) is calculated for noun categorization to four categories and for adjective categorization to two categories. The Kappa scores for two raters are respectively 0.82 and 0.87, both indicating proof of good interrater agreement (Randolph, 2005).

<Table 4>

For multi-word product identification, the four noun and two adjective category labels are applied to the retrieved Part-Of-Speech sequences as follows:

- If the last word is categorized as non-relevant (Noun4) or as an explaining noun (Noun3), the full POS sequence is ignored.

- If the last word is a product by itself (Noun1), the full POS sequence is labelled as product.
- If the last word is a noun that needs explanation (Noun2) and this explanation is given by either an explaining noun (Noun3), product noun (Noun1) or a relevant adjective (Adj1), the Part-Of-Speech sequence is retained as product.

A final processing step lemmatizes the last words of the retrieved multi-word products (e.g. transforming *magnetic strips* to *magnetic strip*) and the frequencies of these lemmatized multi-word products in patents are recorded in a product-patent matrix.

3.5.2 Organism identification

Organism identification is performed in the same manner as for mention filtering in Section 3.2.1. The algorithm detects both single- (e.g. Lotus) and multi-word organism names (e.g. Lotus japonicus), as well as common (e.g. gryfalcon) and scientific names (e.g. Falco rusticolus), and is applied to the titles of the biological documents to record organism occurrences in an organism-paper matrix (Vandevenne et al., 2014b).

3.5.3 Positioning products and organisms in PA and OA space

Following the folding-in procedure (Deerwester et al., 1990), the product-patent and organism-paper matrices are multiplied with respectively the patent-PA and paper-OA matrices calculated in Section 3.3 to obtain two characterization matrices: product-PA and organism-OA. Figure 6 illustrates organism characterization for *geckos* in a radar plot. The visualization shows the most important OAs (highest organism loadings in OA space, obtained from the organism-OA matrix) for the selected organism. This characterization is a reflection of the eight documents currently present in the database that contain a reference to geckos in their title. In general, the more documents associated to an organism in the database, the more the radar plot becomes spiral-shaped as more topics can be addressed. If there is only one document stored for a specific organism, the radar plot typically spikes for one or a few Organism Aspects.

<Figure 6>

4 SEABIRD Functionalities

The algorithms described in the previous section are used to implement the back-end architecture of SEABIRD, as shown in Figure 1. To allow users to interact with the system while moving through the different Biologically-Inspired Design process phases, a number of front-end or user interface elements have been developed, which are detailed below.

4.1 Network Browser

The information structure created with the algorithms described in Section 3, in fact, represents a large network containing four types of nodes: products, Product Aspects, Organism Aspects and organisms. Links are formed between these nodes by consulting the characterization and mapping matrices, i.e. the product-PA, PA-OA and organism-OA matrices. This way, products are connected to PAs, PAs to OAs, and OAs to organisms. A central user interface element, named *network browser*, is developed to assist navigation through this network. The network browser contains four columns, from left to right: products, PAs, OAs and organisms. Table 5 illustrates with an example where the product *display* (column 1) is characterized with Product Aspects (column 2). The selected PA *colour perception* maps to Organism Aspects (column 3) from which the OA *Colour* results in the listed organisms (column 4). For the problem-oriented bio-inspired design process (Helms et al., 2009), the user navigates from left to right in the network browser.

<Table 5>

4.2 Supporting Problem Formulation: BID Process Phase 1

Problems are not rarely formulated as "*How can we improve our product?*". In the context of BID, this question translates to "*How can nature inspire us to improve our product?*". To capture this problem formulation, SEABIRD has a search functionality to select one or more products from the large product database (see Section 3.5.1). First, a direct search allows searching for multi-word products. One can search for the exact product or for products

starting, ending or containing a specified character string. The results, which easily contain tens to hundreds of multi-word products, are visualized as a word cloud, where the font size reflects the importance (frequency) of the products in the patent database. For example, the most important products for *bag* are *bag*, *air bag*, *plastic bag*, *packaging bag*, *golf bag*, etc. These results also contain product parts: for example, when searching for *air bag*, *air bag cover* and *air bag gas* are also found. A second, indirect, search finds products that are mentioned in the same patents as a selected product. Continuing the same example, the most important products linked to *air bag* with indirect search are *vehicle*, *steering wheel*, *gas generator*, *door*, *fabric*, *seat*, etc. This list contains product synonyms (e.g. air bag apparatus), elements of the super system (e.g. vehicle) and elements of the subsystem (e.g. fabric). As this indirect search also has the potential to generate hundreds of product names, again a frequency-weighted word cloud is used as visualization method. Product search and selection is an important step in problem formulation, as different but related products are likely to offer distinct characterizations. For example, the product *car body* is characterized with the PA *aerodynamic*, but, if one would select the product *car* this PA is suppressed by other PAs. Indirect search facilitates the identification of the most relevant product. Product selection adds it to the first column of the network browser and SEABIRD characterizes the product with Product Aspects, as explained in Section 3.5.3). Problem definition ends with the user selecting the PA that best expresses the desired function or aspect of the product to improve.

Another possible way to formulate a problem is "*How can we realise a specific functionality?*" or "*How can nature inspire us to realise a specific functionality?*". As there is no product to start from, a direct selection is made from the set of Product Aspects. For example, for the question "*How can we realize acceleration?*", the PA *acceleration and deceleration* can be selected. SEABIRD has a search function for PAs that, like for products, can search for exact PAs or for PAs starting, ending or containing a specified character string.

For example, searching for *heat* results in the following two PAs: *heating / cooling* and *heat treatment*. These search results are positioned in the second column of the network browser. Independently of which problem formulation strategy is taken, the output of problem formulation is a set of PAs from which the user selects one or more to continue to the next BID process step, explained in the following section.

4.3 Supporting Scalable Search for Biological Strategy Documents: BID Process Phase 2

For each selected Product Aspect, an ordered list of relevant Organism Aspects can be generated in the third column of the network browser. To obtain this list, the PA-OA similarity or mapping matrix is consulted, as explained in Section 3.4. Continuing the running example, the most important PA for *air bag* is *inflation / deflation* and the most relevant OA that maps to this PA is *volume (inflation / deflation)*.

4.3.1 Search for relevant organisms

After selecting an OA in the network browser, its fourth column of the network browser is populated with an ordered list of organism names that reflects the organism loadings stored in the organism-OA matrix. The organisms put forward for the OA *volume (inflation / deflation)* are *mute swan*, *walking goby*, *American grasshopper*, *locusta*, *rana nicobariensis*, *nicobar island frog*, *ribbed mussel*, *bearded seal*, etc. Immediate understanding of which strategies these organisms implement that are relevant for the selected OA is not evident for this example, and further analysis is necessary in the following BID process steps. A second example, generating a list of organisms for the OA *bioluminescence* is easier to directly verify, as the retrieved organisms are: *glowworms*, *Lampyris noctiluca*, *common*, *dinoflagellates*, *Suberites domuncula*, *fireflies*, *brittle stars*, *Photuris*, etc.

4.3.2 Search for relevant strategy documents

For a selected organism aspect, SEABIRD consults the OA-document matrix to identify the most relevant strategy documents with the highest OA loadings. A *strategy document explorer* is developed for SEABIRD that presents the information in table format. Each line contains the strategy title and involved organism name for a quick overview of the results. Additionally, by clicking on a strategy document line the biological paper's abstract is shown together with a link to the full text, for further analysis in the next BID steps. Continuing the running example, high-scoring strategy documents for the OA *volume (inflation / deflation)* discuss volume changes of the lung, the swim bladder and individual cells.

4.4 Supporting filtering: BID process phase 3

With scalable search allowing the integration of very large biological knowledge repositories, a new challenge emerges. The more strategy documents SEABIRD integrates, the more relevant results become retrievable and measures need to be taken to avoid information overload. The presentation of ordered lists of organisms and strategy documents is a first important filtering method. In this way, the user can retrieve as many ordered results as he or she wants to process. A second functionality that deals with growing database size is the organism oriented view that SEABIRD provides. Besides organism characterization, it also presents all strategy documents for the specific organism in a separate table. This way, related studies about similar biological strategies of the same organism are grouped. For example, the organism view contains the characterization as illustrated by Figure 6, but also presents a strategy document table containing eight entries for geckos. Seven of these strategy documents discuss the strategy of gecko adhesion in the context of locomotion, climbing or static hanging. One strategy document focusses on high-frequency gecko communication. As database expansion will pose new challenges, extra filtering measures are considered, as detailed in Section 6, Discussion and future work.

4.5 Supporting analysis and knowledge transfer: BID process phases 3 and 4

After having narrowed down the set of potential biological strategy documents to a manageable number, the retrieved biological knowledge needs to be analysed and transferred to the technical domain by identifying cross-domain analogies. These are cognitive processes. The main information unit that SEABIRD returns are academic papers, which, intuitively, is a too bulky representation to use as stimuli for knowledge transfer. Therefore, initially only the title and abstract are shown with text-highlights for the high-loading terms of the selected OA. Term loadings are retrieved by consulting the term-OA matrix obtained from OAs generation (see Section 3.3). Figure 7 illustrates annotation for an example title and abstract of a strategy document scoring on the OA *adhesion*. The highest loading terms for this OA are *adhesive*, *attach*, *glue*, *detach*, *etc.* These terms and their inflections (e.g. adhesives, adhesiveness, adhesion, adhesions for adhesive) are marked in the text to catch attention. The same annotation process can be applied to the full text if the user requires more background information. Automated text annotation, as described above, builds on the assumption that the terms scoring high on the relevant OAs are interesting to focus on during knowledge transfer.

<Figure 7>

5 Validation of Scalable Search

The main contribution of this paper is presenting an approach for scalable cross-domain search. Therefore the authors validate SEABIRD's implementation of the first and second BID process phase. In fact, after searching with SEABIRD, other research contributions can be integrated to complete the BID process phases. For example, the integration of automated extraction of causally related functions from natural-language text (Cheong & Shu, 2014) could be considered to facilitate knowledge transfer. More opportunities for combining research efforts are discussed in Section 6.

In order to objectively validate the search functionality, a set of search questions and an accompanying set of desirable results are required. This translates to identifying a set of existing bio-inspired concepts and testing if SEABIRD's problem formulation and search functionalities are able to return relevant stimuli for knowledge transfer. For the selection of validation cases, well-known bio-inspired designs are taken, not rarely used as examples for introducing BID, e.g. the butterfly-inspired Mirasol display technology (Mirasol, 2009). By taking widely accepted examples of BID as validation cases, the authors aim at (1) maximizing the likelihood of readers being familiar with the presented validation cases and (2) at minimizing doubt whether the bio-inspired design was really inspired by nature.

As explained in Section 3.1, SEABIRD currently integrates a relatively small database compared to human's knowledge about nature. This entails that the biological strategies relevant for the chosen validation cases are possibly not represented in this corpus by one or more strategy documents. In such cases one or more relevant biological papers are added to SEABIRD, and automatically positioned in the Organism Aspect space with folding-in (Deerwester et al., 1990). Adding biological strategy documents this way does not impede validation as it is not database completeness but the search functionality that is verified. At the same time folding-in extra biological strategy documents illustrates that corpus expansion is possible for SEABIRD without the need to repeat any of the interactive labelling or classification tasks detailed in Section 3 and without the need to regenerate the Organism Aspects.

5.1 Validation case 1: Butterflies and Mirasol display technology

For the development of a new display technology, power consumption is an important parameter. When resource efficiency is a primary goal, nature is an interesting source of inspiration (Bar-Cohen, 2006). The Mirasol display technology (Mirasol, 2009) is inspired by the structural colouring of butterflies. "Structural colour utilizes the wave-nature of light. As a

wave, light can experience wave superposition; that is, groups of waves may add together to reinforce or diminish their combined effect. For this to happen effectively and therefore to produce a distinct colour effect, there must be a definite structural order in the system; importantly, the physical dimension of this order, the period, must be on a par with the wavelength of light. This phenomenon is often referred to as interference and is identical to the mechanism that produces the iridescent colours in soap bubbles; other names for it include Bragg diffraction or coherent scattering" (Vukusic, 2006). Besides power savings, a second competitive advantage of this technology is sunlight viewability (Mirasol, 2009). For the butterfly-inspired displays, the structure is created with microscopic machines that consist of different material layers and a variable size air gap manipulatable with applied voltage (Mirasol, 2009).

SEABIRD's multi-word product search is used to initiate problem formulation. The product word cloud returns *display* as the most important product in the patent corpus for the search key *display*. Other high-ranked results are *display device*, *display screen*, *display apparatus*, *etc.* *Display* is taken as the product to characterize, which results in the following ordered Product Aspects: *text and illustration*, *data transmission*, *illumination*, *colour perception*, *vibrating*, *etc.* The other high-ranking product synonyms are confirmed to have similar characterizations, mainly differing in the order of the Product Aspects. For the problem at hand, the PA *colour perception* is chosen to map to OAs. SEABIRD's PA-to-OA-mapping suggests in decreasing order of importance: *colour*, *colouration*, *spectrum and reflectance*, *spectral*, *optical*, *illumination*, *camouflage*, *etc* and the first OA *colour* is selected to explore for relevant organisms. A list of organisms is generated which, in order of relevance to the selected OA, are: *swallowtails*, *strawberry poison frog*, *Steller's jay*, *monarch butterfly*, *cephalotes atratus*, *Graphium sarpedon*, *budgerigar*, *butterflies*, *Nymphalini*, *eastern nosquitifish*, *etc.* Based on SEABIRD's current corpus, five of the top ten results are a member

of the superfamily *Papilionoidea* with common name *butterflies*. The Linnaeus classification of the identified *Papilionoidea* ranges from the species level (monarch butterfly) to the super family (butterflies). Next, the SEABIRD strategy document explorer's results are scanned to verify the presence for one or more specific biological strategies that have the potential to trigger the design of the Mirasol display technology. For the OA *colour*, the eleventh strategy document is titled "Significance of a basal melanin layer to production of non-iridescent structural plumage colour: evidence from an amelanotic Stellers jay", discussing structural colouration for birds. For the OA *colouration*, the ninth result is titled "Spectral reflectance and directional properties of structural colouration in bird plumage" and for the OA *spectrum/reflectance* the second, eight and fourteenth result are respectively titled "Glass scales on the wing of the swordtail butterfly *Graphium sarpedon* act as thin film polarizing reflectors", "Blue integumentary structural colours in dragonflies (Odonata) are not produced by incoherent Tyndall scattering" and "Anatomically diverse butterfly scales all produce structural colours by coherent scattering". Each of these titles describe a biological strategy strongly related to the Mirasol technology. Therefore, for this validation case, the search results can be said to provide useful strategy documents as input to the fourth BID process step, knowledge transfer, to lead to the conceptual design of the successful display technology.

As illustrated above, SEABIRD makes it possible to identify similar solutions in nature for taxonomically distant species. Structural colouring is not a strategy exclusive to butterflies: "The non-pigmentary source of colour, referred to as structural colour, is a very important component in the appearance of many different animal systems; examples are found in many other orders of insects, as well as in birds and aquatic animals." (Vukusic, 2006). Indeed, SEABIRD, only relying on the biological papers of one journal, retrieves two strategy

documents describing structural colouring for birds, two for butterflies and one for dragonflies.

5.2 Validation case 2: Boxfishes and the Mercedes-Benz Bionic concept car

A key focus point in new car development is energy consumption in which aerodynamics plays an important role. During product search with SEABIRD, as a part of problem formulation, it soon becomes clear that there are a large number of multi-word products in the patent database containing the word *car*. The most important are *car*, *elevator car*, *electric car*, *railway car*, *motor car*, *etc.* Hence, product search is focussed on the part of the car most relevant to aerodynamics, being the *car body*. This product is found with indirect product search, see Section 4.2, locating products that frequently co-occur with the product *car*. SEABIRD's characterisation of *car body* places the PA *aerodynamic* on the second place, right after *rigidity/deformation/elasticity*. Mapping to OAs identifies the OA *vortex* on the fourth place and the retrieved organisms are in order of relevance: *mullets*, *swimming frogs*, *boxfishes*, *Sarsia tubulosa*, *thrushes*, *mayflies*, *blackcap*, *swift*, *etc.* The strategy document explorer finds one relevant strategy document discussing boxfishes for the OA *vortex* titled: "Body-induced vortical flows: a common mechanism for self-corrective trimming control in boxfishes". In this study, "flows around the bodies of three morphologically distinct boxfishes" are investigated and it was found that "carapaces of boxfishes, which vary in cross-sectional shape, longitudinal features and ornamentation, play an important role in hydrodynamic stability" (Bartol et al., 2005). Hence, SEABIRD's problem formulation and scalable search have made the link from aerodynamics of the car body to the shape of the body of boxfishes.

5.3 Validation case 3: Termite mound and air conditioning for buildings

Temperature regulation in large office buildings is conventionally realized with air-conditioning and heating systems. The Eastgate Centre in Harare, Zimbabwe, has a

temperature regulation system inspired by self-cooling termite mounds. At the time of inspiration¹, the biological understanding of temperature control in termite mounds had two main components: (1) less dense, warm internal air rises and is exchanged in the chimneys with denser cooler air that moves downward; (2) an induced-flow mechanism (Venturi effect) causes air to enter through the cavities near the ground (lower wind velocity) and to exit through the chimneys (higher wind velocity). Application of this strategy causes the Eastgate centre to save up to 90% in energy costs.

Unlike the previous two examples, SEABIRD's current corpus did not contain any strategy document that could function as valid search result. As explained in Section 6, in such a case the corpus is extended with one or more relevant strategy documents to validate the search functionality. As a bonus, at the same time the folding-in process for corpus expansion is tested. For this validation case, four biological strategy documents were folded-in OA space (see Table 6). The second strategy document in this list does not mention an organism name, hence it will not contribute to the organism lists generated for OAs. It can, however, be retrieved via the strategy document browser as this functionality does not require an identified organism in the title.

<Table 6>

SEABIRD's multi-word product search returns a number of candidate products, e.g. *air conditioning system*, *air conditioning apparatus*, *air conditioning device*. etc. The most important Product Aspect for these products is *heating/cooling/temperature*. Mapping this PA to OAs results in the OA *temperature* on the fourth place and *termites* are placed fourth in the organism list for this OA. This can motivate the user to explore all strategy documents of termites, which returns the three validation strategy documents linked to termites and two more already present in the corpus about unrelated topics. The strategy document ranked first

¹ Today, a more complex understanding of how termite mounds work is reported (Turner & Soar, 2008). This however does not change the fact that the Eastgate Centre was bio-inspired, and the documents detailing the biological strategies remain valid stimuli that can trigger successful BID designs.

by SEABIRD's strategy browser is "Nest Thermoregulation in Social Insects"; and the ranks of the other strategy documents are detailed in Table 6. Hence the folded-in strategy documents are retrievable by SEABIRD's search functionality. So again for this case, SEABIRD's search is able to retrieve stimuli that have the potential to trigger a bio-inspired invention.

6 Discussion and Future Work

Although eventually SEABIRD aims at effectively supporting the designer in all four SSBID phases, in this paper, the contribution of adding scalability to the search phase is the main focus. Therefore SEABIRD's search is validated with three well-known BID cases. For each case the proposed system was able to identify relevant stimuli with the potential to lead to the development of bio-inspired innovations. The confidence about SEABIRD's search functionality obtained from these validation tests, combined with extensive corpus expansion and further development of algorithms to support the problem formulation, filtering, analysis and knowledge transfer phases, will allow testing the ideation tool in controlled outcome-based experiments were new design challenges are presented to a large group of participants.

Table 1 indicates that, compared to related research, SEABIRD's biological strategy corpus is large. However, compared to the end goal of leveraging human's knowledge about nature, significant corpus expansion is required. SEABIRD sets itself apart by its inherent scalability, eliminating any limit for increasing the supporting biological database while still facilitating efficient search. The interactive operations reported in Section 3—the term filtering and categorization steps—take place during system setup and need to occur only once for each unique term in the corpus dictionary that survived the automated term filtering steps. These tasks, requiring human interaction, take together maximum five days. Adding new documents to the corpus gradually has less and less influence on dictionary size—that is, fewer new

terms are introduced to a growing dictionary by extra documents. Although the stability of the current dictionary is not yet claimed in this contribution, the current corpus and corpus dictionary already allowed validation of the presented BID cases. Future work will include a second iteration for SEABIRD's setup to increase the corpus until the critical mass is reached that delivers a stable dictionary. Thereafter, no more interactive labeling or categorization tasks are required to integrate as many new sources as there are—or become—available. Of course, after a number of years the dictionary should be evaluated as language itself evolves. If necessary, a small effort can integrate new terms. A similar reasoning holds for the interactive labeling of Organism Aspects. By comparing the Organism Aspect set for a gradually increasing database, a minimal corpus size will be determined where the Organism Aspect set becomes stable, as previously illustrated for the Product Aspect set of the PAnDA tool (Verhaegen et al., 2011a). Thereafter, more biological strategy documents will be folded-in (Deerwester et al., 1990) in OA space without any human interaction besides selecting the corpus and running the scripts. Compared to the state-of-the-art, the above interactive labeling or classification tasks occur on the corpus level not the document level, hence they only need to be performed once for a stable corpus and dictionary. As these tasks require a relatively small human effort during SEABIRD's setup, they are not the focus of future automation attempts.

The above listed interactive tasks during SEABIRD's initial setup require human judgment and are, hence, subjective and even error-prone. Although the confirmed repeatability gives some confidence about the execution of these tasks, they also indicate that opinions for term labeling or classification can differ. The effect of this, however, is limited due to the key role of dimension reduction (PCA) in SEABIRD's design. This technique combines correlated original variables (term frequencies) to uncorrelated variables (PAs and OAs). Take, for

example, the OA depicted in Figure 5. If one of these terms (correlated variables) would be mislabeled, there would still be a set of correlated variables representing a concept in the data for PCA to reveal. As the corpus dictionary grows, the potential number of terms behind the concepts grows and the technique becomes more and more resistant to such errors. Of course there is a limit to this appealing property, but the validation cases illustrate that the execution of the interactive, repeatable tasks on a corpus of 8011 documents already results in a functional bio-ideation tool. A further illustration of the resistance of the functioning of the proposed tool to changes in the retained dictionary is the fact that only the terms that occur at least ten times in the corpus are interactively evaluated, the rest does not take part in dimension reduction. The central role of semantic concepts linking the technical to the biological domain during search—as opposed to direct search with keywords—has the advantage that the identification of cross-domain links is not dependent of finding the right keywords, but of choosing the right concepts from the offered product characterization.

With further corpus expansion, the need for more support in search result filtering and grouping will increase (SSBID phase 3). Therefore, measures like clustering related strategy documents discussing the same strategy, biological scale detection and taxonomic visualization of search results are envisaged. Related strategies could, for instance, also be grouped by detecting enabling functions (Cheong & Shu, 2014).

Further expansion of the PA and OA sets will increase the ability to specify a technical problem and to find matches with OAs on the conceptual level. For example, there is an OA *adhesion* that scores high on the characterization of many organisms like *geckos*, *carnivorous plants*, *spiders*, *mussels*, *etc.*; and in the patent database many products that have a functionality related to adhesion or attachment are identified (fastener tape, adhesive tape, paper glue, fastener, *etc.*). However, the most relevant concept in the set of 300 PAs for attachment functionality currently is *soldering*, clearly linked to a subset of products while a

semantically more general concept currently is missing. This example illustrates the potential benefit of generating more Product and Organism Aspects than the currently arbitrarily chosen sets of 300. Another feature that would benefit SEABIRD is the ability of multiple selection of products, Product Aspects or Organism Aspects. This way synonymous products (e.g. computer screen and computer display) could be characterized as one, or multiple Product Aspects could be combined to map to Organisms Aspects, or multiple Organism Aspects could be selected to generate relevant organism or strategy document lists.

In this contribution the authors have split the validation of access to relevant biological strategies from testing how the natural language stimuli should be represented for efficient ideation by facilitating the recognition of cross-domain analogies. The former is an Information Retrieval problem, the latter a cognitive science problem which the authors hope to address in future research. For the information retrieval problem, it is validated that SEABIRD offers relevant stimuli via a logical path. How many users would indeed recognize these solutions is not yet tested as this task requires solving the second important challenge of effective stimuli representation for supporting knowledge transfer. This task—identifying valid analogies in natural-language text resulting from search—is not trivial (Cheong & Shu, 2013), even if relevant biological text is presented. Section 4.5 details current efforts to support knowledge transfer and related contributions suggest more measures can be taken to increase SEABIRD's support for this BID process phase. These range from training users in Biologically-Inspired Design to adding extra knowledge transfer supporting functionality to the ideation tool. Nelson et al. (2009) found that training BID students helps them to develop more novel and more diverse design ideas in a test setup without biological strategies as stimuli. To counteract cognitive bias, the abstraction of nouns to their hypernyms in biological texts is proposed (Cheong & Shu, 2013), a process close to automation; and another contribution identifies causally related verbs (Cheong & Shu, 2014) to support structural

mapping (Gentner, 1997). Furthermore, it is likely that instantiating knowledge transfer models (Functional Basis, Structure Behaviour Function and SAPPhIRE) for both the biological systems as for the formulated technical problem will benefit the recognition of cross-domain design-by-analogy opportunities. For SBF models, a pilot study indicates that some designers benefit from SBF model instantiation for understanding biological articles (Vattam & Goel, 2011). Although manual model instantiation does not scale well for a large biological database, this is an opportunity to leverage these academic insights in knowledge transfer when a limited set of relevant biological strategies have been identified by SEABIRD.

7. Conclusion

To eliminate the element of chance in discovering new bio-inspired solutions to technical problems, a number of Systematic BID approaches have been proposed. However, all approaches struggle with integrating the ever-growing body of biological knowledge in a scalable way. Therefore, Scalable sEarch for systemAtic Biologically-InspiRed Design (SEABIRD) is proposed, a bio-ideation methodology that leverages large natural-language biological databases in the search for relevant biological stimuli for design-by-analogy.

Central to SEABIRD are Product and Organism Aspects, two concept sets extracted from respectively a patent and biological paper database. Problem formulation is supported by combining an advanced multi-word product search with product characterization, resulting in one or more Product Aspects to focus on. A mapping of Product to Organism Aspects results in ordered strategy document and organism lists, which are initial stimuli for the filter analysis and knowledge transfer phases.

SEABIRD's search functionality is validated by demonstrating its capability to provide relevant stimuli with the potential to inspire three well-known BID cases. During validation,

SEABIRD's ability to easily expand the biological corpus is also illustrated. Indeed, no interactive tasks, that have been found to encumber the scalability of existing systematic BID approaches, are required during corpus expansion. After demonstrating a novel approach for scalable search, focus needs to be shifted to increase support to the other SSBID phases. By combining findings from existing research with the development of extra supporting functionalities, SEABIRD needs to be extended to effectively support bio-inspired concept generation.

References

- Altshuller, G.S. (1984). *Creativity as an Exact Science: The Theory of the Solution of Inventive Problems*. Gordon and Breach Science Publishers.
- Bajželj, B., Allwood J.M., & Cullen J.M. (2013). Designing Climate Change Mitigation Plans That Add Up. *Environ. Sci. Technol.* 47(14), 8062-8069.
- Bar-Cohen, Y. (2006). Biomimetics--using nature to inspire human innovation. *Bioinspiration & Biomimetics* 1, 1-12.
- Bar-Cohen, Y. (2011). *Biomimetics: Nature-Based Innovation*. CRC/Taylor & Francis.
- Bartol, I.K., Gharib, M., Webb, P.W., Weihs, D., & Gordon, M.S. (2005). Body-induced vortical flows: a common mechanism for self-corrective trimming control in boxfishes. *The journal of experimental biology* 208, 327-344.
- Benyus, J.M. (1997). *Biomimicry: Innovation Inspired by Nature*. Harper Perennial.
- Berry, M.W., Dumais, S.T., & O'Brien, G.W. (1995). Using Linear Algebra for Intelligent Information Retrieval. *SIAM Review* 37(4), 573-595.
- Bhushan, B. (2009). Biomimetics: lessons from nature—an overview. *Phil. Trans. R. Soc. A* 367(1893), 1445-1486.
- Bonser, R.H.C. (2006) Patented Biologically-Inspired Technological Innovations: A twenty year review. *Journal of bionic engineering* 3, 39-41.
- Bonser, R.H.C., & Vincent, J.F.V. (2007). Technology trajectories, innovation, and the growth of biomimetics. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 221(10), 1177-1180.

- Brants, T. (2000). TnT: a statistical part-of-speech tagger. *Proceedings of the sixth conference on Applied natural language processing*, ANLC '00, 224-331.
- Chakrabarti, A., Sarkar, P., Leelavathamma, B., & Nataraju, B.S. (2005). A functional representation for aiding biomimetic and artificial inspiration of new ideas. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 19(2), 113-132.
- Charniak, E. (1997) Statistical Techniques for Natural Language Parsing, *AI Magazine* 18(4), 33-43.
- Cheong, H., Chiu, I., Shu, L.H., Stone, R.B., McAdams, D.A. (2011) Biologically Meaningful Keywords for Functional Terms of the Functional Basis. *Journal of Mechanical Design* 133(2), 021007.
- Cheong, H., Hallihan, G., & Shu, L.H. (2012a) Understanding analogical reasoning in biomimetic design: an inductive approach. *Design computing and cognition DCC'12*.
- Cheong, H, Shu, L.H. (2013). Reducing cognitive bias in biomimetic design by abstracting nouns. *CIRP Annals - Manufacturing Technology* 62(1), 111-114.
- Cheong, H, Shu, L.H. (2014). A Method to Retrieve Causally Related Functions from Natural-language Text for Biomimetic Design, *Journal of Mechanical Design*, DOI: 10.1115/1.4027494.
- Chiu, I. and Shu, L.H. (2005). Bridging cross-domain terminology for biomimetic design. *Proc. of the ASME 2005 International Design Engineering Technical Conferences*, DETC2005-84908.
- Chiu, I., & Shu, L.H. (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.
- Deerwester, S.C., Dumais, S.T., Landauer, T.K., Furnas, G.W., & Harshman, R.A. (1990). Indexing by latent semantic analysis. *Journal of the American Society of Information Science* 41(6), 391-407.

Deldin, J.M., & Schuknecht M. (2014) The AskNature Database: Enabling Solutions in Biomimetic Design. In *Biologically Inspired Design* (Goel, A.K. et al. eds.), pp. 17-27. DOI: 10.1007/978-1-4471-5248-4_2

Fall, C.J., Törösvári, A., Benzineb, K., & Karetka, G. (2003) Automated Categorization in the International Patent Classification. *ACM Special Interest Group on Information Retrieval Forum* 37(1), 10-25.

Fellbaum, C. (1998). WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press.

Fox, C. (1998). A stop list for general text. *ACM Special Interest Group on Information Retrieval Forum* 24(1-2), 19-21.

Gebeshuber, I.C., Gruber, P., & Drack, M. (2009). A gaze into the crystal ball: biomimetics in the year 2059. *J. Mechanical Engineering Science* 223(12), 2899-2918.

Gentner, D. en Markman, A.B. (1997). Structure Mapping in Analogy and Similarity. *American Psychologist*. 52(1), 45-56.

Gerner, M., Nenadic, G., & Bergman, C. M. (2010). LINNAEUS: a species name identification system for biomedical literature. *BMC bioinformatics* 11(85). DOI:10.1186/1471-2105-11-85

Goel, A.K., Vattam, S., Wiltgen, B. & Helms M. (2012). Cognitive, collaborative, conceptual and creative. Four characteristics of the next generation of knowledge-based CAD systems: A study in biologically inspired design. *Computer-Aided Design* 44(10), 879-900.

Google Merchant Center (2014). Categorize your products, available at <https://support.google.com/merchants/answer/160081?hl=en> (accessed 14.05.08).

Helms, M., Vattam, S.S., & Goel, A.K. (2009). Biologically inspired design: process and products. *Design Studies* 30, 606-622.

IPCC, Intergovernmental Panel on Climate Change (2007) IPCC Fourth Assessment Report Working Group 1 report. ISBN: 978-0-521-88009-1.

Kaiser, H.F. (1958). The Varimax criterion for analytic rotation in factor analysis. *Psychometrika* 23(3), 187-200.

Kaiser, M. K., Farzaneh, H.H. and Lindemann, U. (2012) An approach to support searching for biomimetic solutions based on system characteristics and its environmental interactions. Proc. of Design 2012, 969-978. Kaiser, M. K., Farzaneh, H.H. and Lindemann, U. (2014) BIOscrabble – The role of different types of search terms when searching for biological inspiration in biological research articles. Proc. of Design 2014, 241-250.

Landauer, T.K. and Dumais, S.T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review* 104(2), 211-240.

Larkey, L.S. (1999) A patent search and classification system. *Proceedings of the fourth ACM conference on Digital libraries*, New York, NY: ACM Press, 179-187.

Lenau, T., Dentel, A., Ingvarsdóttir, Þ., & Guðlaugsson, T. (2010). Engineering design of an adaptive leg prosthesis using biological principles. *Proc. of DESIGN 2010*, 331-340.

Lepora, N.F., Verschure, P., & Prescott, T.J. (2013). The state of the art in biomimetics. *Bioinspiration & Biomimetics* 8(1).

Linnaeus, C. (1767). *Systema Naturae*.

Mak, T.W., & Shu, L.H. (2008). Using descriptions of biological phenomena for idea generation. *Research in Engineering Design* 19(1), 21-28.

- Mirasol (2009). Competitive display technologies (White paper). Available at: <http://www.qualcomm.com/sites/default/files/uploads/competitivedisplaytechnologies-06-2009.pdf>
- Murphy, J.T. (2011). Patent-based analogy search tool for innovative concept generation. PhD thesis, Department of Mechanical Engineering, The University of Texas at Austin.
- Nagel, J.K.S., Nagel, B.I., Stone, R.B., & McAdams D.A. (2010). Function-based, biologically inspired concept generation. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 24(4), 521-535.
- Nagel, J.K.S., & Stone R.B. (2012). A computational approach to biologically inspired design. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 26(2), 161-176.
- Nelson, B., Wilson, J., & Yen J. (2009). A study of biologically-inspired design as a context for enhancing student innovation. *39th ASEE/IEEE Frontiers in Education Conference*, T2F 1-5.
- Purves, W. K., Sadava, D., Orians, G. H., & Heller, H. C. (2001). *Life, The Science of Biology*, 6th ed., Sinauer Associates, Sunderland, MA.
- Randolph, J.J. (2005). Free-marginal multirater kappa: an alternative to Fleiss' fixed-marginal multirater kappa. *Joensuu University Learning and Instruction Symposium, Joensuu, Finland*.
- Salton, G., Wong, A., & Yang, C.S. (1975). A Vector Space Model for Automatic Indexing. *Communications of the ACM* 18(11), 613-620.
- Salton, G. & Buckley, C. (1988) Term-weighting approaches in automated text retrieval. *Information Processing & Management* 24(5), 513-523.
- Sartori, J., Pal, U., & Chakrabarti, A. (2010). A methodology for supporting "transfer" in biomimetic design. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 24(4), 483-505.

Skillicorn, D. (2007). Understanding Complex Datasets: Data Mining with Matrix Decompositions. Chapman & Hall / CRC Data Mining and Knowledge Discovery Series. ISBN: 978-1584888321.

Shu, L.H. (2010). A natural-language approach to biomimetic design. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 24(4), 507-519.

Srinivasan, V., Chakrabarti, A., & Lindemann, U. (2012). Towards an ontology of engineering design using SAPPhIRE model. *Proc. of CIRP Design 2012*.

Vandevenne, D., Verhaegen, P., Dewulf, S., & Duflou, J.R. (2011). A scalable approach for the integration of large knowledge repositories in the Biologically-Inspired Design process. *Proc. of the 18th International Conference on Engineering Design (ICED11)* 6, 210-219.

Vandevenne, D., Kellens, K., Banck M., Clijsters, H., Verhaegen, P.-A., & Duflou, J.R. (2012a). A framework for assessing the Sustainability of Biologically-Inspired Products. *Innovation for Sustainable Production, i-sup*.

Vandevenne, D., Verhaegen, P.-A., Dewulf, S., & Duflou, J.R. (2012b). Automatically populating the Biomimicry Taxonomy for scalable systematic Biologically-Inspired Design. *Proc. of IDETC2012*. DOI: 10.1115/DETC2012-70928.

Vandevenne, D., Verhaegen, P.-A., Dewulf, S., & Duflou, J.R. (2014a). A scalable approach for ideation in biologically inspired design. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, DOI: 10.1017/S0890060414000122.

Vandevenne, D., Verhaegen, P.-A. & Duflou, J.R. (2014b). Mention and focus organism detection and their applications for scalable systematic bio-ideation tools. *Journal of Mechanical Design*, expected publication November 2014.

Vandevenne, D. & Duflou, J.R. (2014c). Methods and Algorithms for Systematic Biologically-Inspired Design. PhD thesis, KU Leuven, ISBN: will be available during typesetting.

- Vattam, S.S., Helms, M.E., & Goel, A.K. (2010a). A content account of creative analogies in biologically inspired design. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 24(4), 467-481.
- Vattam, S, Wiltgen, B., Helms, M., Goel, A.K., & Yen J. (2010b). DANE: Fostering Creativity in and through Biologically Inspired Design. *Proc. of International Conference on Design Creativity*, 115-122.
- Vattam, S.S., & Goel, A.K. (2011). Semantically annotating research articles for interdisciplinary design. *K-CAP'11*, 165-166.
- Verhaegen, P.-A., D'hondt, J., Vertommen, J., Dewulf, S., & Duflou, J.R. (2009) Quantifying and Formalizing Product Aspects through Patent Mining. *Proceedings of the ETRIA TRIZ Future 2009*.
- Verhaegen, P.-A., D'hondt, J., Vandevenne, D., Dewulf, S. & Duflou, J.R. (2011a). Identifying Candidates for Design-by-Analogy. *Computers in Industry* 62 (4), 446-459.
- Verhaegen, P., Peeters, J., Vandevenne, D., Dewulf, S., & Duflou, J.R. (2011b). Effectiveness of the PAnDA ideation tool. *Procedia Engineering* 9, 63-76.
- Verhaegen, P.-A., & Duflou, J.R. (2013). Methods and Algorithms for Systematic Innovation. PhD thesis, KU Leuven, ISBN: 978-94-6018-685-1.
- Vincent, J.F.V., Bogatyreva, O.A., Bogatyrev, N.R., Bowyer, A., & Pahl, A.K. (2006) Biomimetics: its practice and theory. *Journal of the Royal Society: Interface* 3(9), 471–482.
- Vukusic, P (2006). Structural colour in Lepidoptera. *Current Biology* 16(16), R621-R623.
- Wilson, J.O., Rosen, D., Nelson, B.A., & Yen, J. (2010) The effects of biological examples in idea generation. *Design Studies* 31(2), 169-186.

Zipf, G.K.(1932). *Selective Studies and the Principle of Relative Frequency in Language*. Harvard University Press.

Table 1: Overview of existing database sizes and content

Method	Size and Content	Reference
Bridge verbs	1 biological introductory handbook	(Shu, 2010)
SBF	40 of which 22 complete SBF models of biological systems	(Vattam et al., 2010)
Functional basis	30 models of biological phenomena	(Nagel et al., 2010)
SAPPhIRE	20 biomimetic examples (engineering and biological systems) 100 biological strategies about motion in nature	(Chakrabarti et al., 2005)
AskNature	1665 detailed descriptions of biological strategies	(Deldin & Schuknecht, 2014)
BioTRIZ	2500 conflicts, from an analysis of 500 biological phenomena	(Vincent et al., 2006)
BIOScrabble	Pubmed, a very large biomedical database ²	(Kaiser et al., 2012)

² This source contains, besides biological strategies, also many articles that are not useful for bio-inspiration (drug testing, mapping genomes, genetically modified organisms, biomedical research methodologies, etc). The authors state: a useful knock-out criterion was, e.g., “the paper is treating a medical application” (Kaiser et al. 2014).

Table 2: Preprocessing example: from corpus document to document vector

Word	POs category	Lemma	Filtered by	Document vector
study	noun	study	relevance filter	1
investigates	verb	investigate	relevance filter	
aerodynamic	adjective	aerodynamic		
gravitational	adjective	gravitational		
forces	noun	force		
ideal	adjective	ideal	relevance filter	
falcons	noun	Falcon	mention filter	
uses	verb	use	stop word filter	
mathematical	adjective	mathematical	relevance filter	
model	noun	model	relevance filter	
calculate	verb	calculate	relevance filter	
speed	noun	speed		
acceleration	noun	acceleration		
diving	noun	diving		

Table 3: The first 20 Organism Aspects

1. concentration	2. diving	3. force	4. oxygen
5. sound	6. olfactory	7. action potential	8. temperature
9. flying	10. swimming	11. magnetic	12. hypoxia
13. spectral	14. walking	15. jumping	16. secretion
17. capture	18. shorten/lengthen	19. current	20. motility

Table 4: Noun and adjective categories

Category	Description	Examples
Noun1	products	valve, display, vehicle
Noun2	potential products, need explanation	system, device, apparatus
Noun3	nouns with explanatory function	temperature, pressure, power
Noun4	all nouns not in Noun1-3	data, portion, end
Adj1	relevant for explaining Noun2	optical, magnetic, digital
Adj2	not relevant for explaining Noun2	preferred, first, improved

Table 5: Representation of the network browser

product	Product Aspect	Organism Aspect	organism
display	text and illustration	colour	swallowtails
	data transmission	colouration	strawberry poison frog
	illumination	spectrum/reflectance	steller's jay
	colour perception	spectral	monarch butterfly
	vibrating	optical	cephalotes atratus

Table 6: Biological strategy documents for air conditioning of buildings, and their rank on the OA *temperature*

Strategy title	Organism	Rank
Ventilation of termite mounds: new results require a new model	termites	60
Nest Thermoregulation in Social Insects	(none)	1
Thermoregulation of termite mounds	Termites	31
Wind-induced ventilation of the giant nests of the leaf-cutting ant Atta vollenweideri	Atta vollenweideri	11

Figure 1: Overview of the proposed system

Figure 2: Preprocessing, from corpus to Document-Term Matrix

Figure 3: Illustration of Zipf's law for the corpus dictionary term frequencies

Figure 4: The Organism Aspect Generation process

Figure 5: Term loadings for an example Organism Aspect

Figure 6: Radar plot of automatic organism characterization of *geckos*

Title: Frictional **adhesion** : A new angle on gecko **attachment** .

Abstract: Directional arrays of branched microscopic setae constitute a dry **adhesive** on the toes of pad-bearing geckos, nature's supreme climbers. Geckos are easily and rapidly able to **detach** their toes as they climb. There are two known mechanisms of **detachment** : (1) on the microscale, the seta **detaches** when the shaft reaches a critical angle with the substrate, and (2) on the macroscale, geckos hyperextend their toes, apparently peeling like tape. This raises the question of how geckos prevent **detachment** while inverted on the ceiling, where body weight should cause toes to peel and setal angles to increase. Geckos use opposing feet and toes while inverted, possibly to maintain shear forces that prevent **detachment** of setae or peeling of toes. If **detachment** occurs by macroscale peeling of toes, the peel angle should monotonically decrease with applied force. (...)

Figure 7: annotation for an example strategy document for the OA *adhesion*

Dennis Vandevenne obtained three ICT-related masters: master in electronics-ICT, master in artificial intelligence-engineering & computer science, and master in industrial management-ICT. From 2006 until present, he is a researcher at KU Leuven. Until 2009 he performed research on identity management (IDM) and biometrics at the Department of Computer Security and Industrial Cryptography (COSIC). Currently he performs research on methods and algorithms for Biologically Inspired Design at the Centre for Industrial Management.

Paul-Armand Verhaegen holds a master degree in applied science and engineering, specialization electrotechnical and computer science, a graduate in the complementary studies in business administration, an MBA from Vrije Universiteit Brussel and a PhD in Engineering from KU Leuven, Belgium. He has founded and sold Stocks, a company specialized in printer supplies. He has worked as an external consultant for 3E on green energy certificates, and as a consultant at Bureau van Dijk Management Consultants. He has worked at Vrije Universiteit Brussel, Erasmushogeschool Brussel and KU Leuven as a research assistant. He is currently working as a post-doctoral researcher in the domain of systematic innovation, product development and data mining.

Simon Dewulf is founder of CREAX. CREAX develops the www.creationsuite.com, a web-based open innovation interface. CREAX Creation Suite is used in companies like Dow Corning, SKF, P&G, Philips, Kraft, L'Oreal, Johnson & Johnson, and GSK to provide a direct access to out of domain knowledge for technology transfer and open innovation opportunities.

Joost R. Duflou holds master degrees in architectural and electro-mechanical engineering and a PhD in engineering from KU Leuven, Belgium. He is a tenure professor in the Department of Mechanical Engineering of this institute. His principal research activities are situated in the field of design support methods and methodologies, with special attention for Systematic Innovation, Ecodesign and Life Cycle Engineering. He is a member of CIRP and has published over 200 international publications.