Artificial Intelligence for Engineering Design, Analysis and Manufacturing (2016), *30*, 78–95.

doi:10.1017/S0890060415000177

SEABIRD: Scalable search for systematic biologically inspired design

DENNIS VANDEVENNE,[1] PAUL-ARMAND VERHAEGEN,[1] SIMON DEWULF,[2] AND JOOST R. DUFLOU[1]

[1]Centre for Industrial Management, Department of Mechanical Engineering, Katholieke Universiteit Leuve, Celestijnenlaan, Leuven, Belgium
[2]AULIVE NV, Ieper, Belgium

(Received May 15, 2014; Accepted December 12, 2014)

Abstract

As more and more people are increasingly turning to nature for design inspiration, tools and methodologies are developed to support the systematic bioideation 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, a scalable search for systematic biologically inspired design (SEABIRD) system is proposed. This system leverages experience from the product aspects in design by analogy tool that identifies candidate products for between-domain design by analogy. 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 bioinspired 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: Bioinspired Design; Biologically Inspired Design; Biomimicry; Creativity and 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 in 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, and sensors. 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 bioinspired 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, and regulating temperature) are similar to the problems humans face. One common rationale motivating BID recognizes the relevance and proven performance of biological solutions (Bar-Cohen, 2006, 2011): why reinvent 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 (Intergovernmental Panel on Climate Change, 2007). Organisms rely on renewable resources for their "production processes," and nature does not generate ever-growing waste piles. Although bioinspired products are not inherently sustainable (Vandevenne, et al., 2012), 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 pos-

Reprint requests to: Dennis Vandevenne, Centre for Industrial Management, Department of Mechanical Engineering, Katholieke Universiteit Leuve, Celestijnenlaan 300A, 3001 Leuven, Belgium. E-mail: dennis.vandevenne@kuleuven.be

sible advantages of bioinspired products are their enhanced marketability and financial savings through efficient use of energy and other resources.

These high expectations of BID 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 bioinspiration for many bioinspired 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 BID 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. About 1.7 million species are named currently, and the total number is estimated to be 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, many sources exist, 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 scalable search for systematic BID (SEABIRD). In Section 2 the state of the art in the systematic search for bioinspiration is listed and discussed. Next, SEABIRD's architecture and functionalities are detailed in Sections 3 and 4, respectively. Thereafter, the proposed approach for scalable search is validated in Section 5 and discussed in Section 6, and conclusions are drawn in Section 7.

2. RELATED RESEARCH

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

2.1. SSBID process phases

A comparison of different contributions describing the encountered phases of the BID process, identifies four ubiquitous phases (Sartori, 2010): formulating search objectives, searching for biological analogues, analyzing biological analogues, and knowledge transfer. These four phases are applied to the SSBID process, which draws inspiration from large biological repositories in a 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 a 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 coming up with a feasible, biologically inspired technical concept.

The above SSBID 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 a 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 because this is where SEABIRD's main contribution currently lies. A more detailed description of these approaches, positioning each in the above four SSBID process phases, is available in Vandevenne, Verhaegen, and Duflou (2014).

2.2. Searching for biological strategies

As a summary of the state of the art in the SBID search phase, this section details three keyword-based search methods, then two approaches supporting on the classification of biological

https://doi.org/10.1017/S0890060415000177 Published online by Cambridge University Press

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, semiautomatic 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 (BIOscrabble) performs a 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 that 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 bioinspiration, 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 bioinspiration need to formulate their design problem in this taxonomy. A classification algorithm was recently proposed that automatically positions biological strategies in the functional categories of AskNature's Biomimicry Taxonomy (Vandevenne, Verhaegen, et al., 2014).

Three approaches can be found in the literature that require human interaction to instantiate detailed models for each biological phenomenon to be integrated in a structured knowledge base. To enable searching, 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 knowledge base generates candidate stimuli for design by analogy. Such a methodology has currently been reported for structure–behavior–function (SBF) models (Vattam, Wiltgen, et al., 2010; 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 cataloging 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 FB terms into biological corresponding terms, is proposed to extend the FB approach. This way, model instantiation, requiring interactive work by humans, for large biological databases is avoided because 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 bioinspired 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.

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

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 & 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[a]	Kaiser et al., 2012

Note: SBF, Structure–behavior–function.

[a]This source contains, besides biological strategies, many articles that are not useful for bioinspiration (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).

https://doi.org/10.1017/S0890060415000177 Published online by Cambridge University Press

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, that is, 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 the following:

- *Interactive result filtering:* For the above natural-language keyword search methods, interactive result relevance filtering does not scale well for large repositories. For the bridge verb approach, it was stated: "Even with a single text used as the source, there can be an unmanageable number of matches" (Shu, 2010). For BIOscrabble, students performed an interactive analysis of 3416 research articles. The large effort analyzing these results concludes that 115 articles, or 3.36%, 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 because they require a detailed analysis of both the engineering and the biological systems to express them on a common abstraction level.
- *Crowdsourcing:* The SBF-based approach developed a social citation cataloging 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 are 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 make 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, Verhaegen, et al., 2014).

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 OA 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.

The extraction of PAs from a patent database is fully described in Verhaegen et al. (2009) and Verhagen, D'hondt, et al. (2011). PAs support the Product Aspects in Design by Analogy (PAnDA) tool that identifies candidate products for design by analogy. Using PAnDA, a product, for example, *carburetor*, can be associated to the most relevant PAs, for example, for the carburetor case *combustion*, *inflow/outflow*, and *rotate*. Based on a selected PA, for example, inflow/outflow, candidate products are depicted as stimuli for the redesign of the carburetor product; for example, 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, Peeters, et al., 2011).

Analogously to the extraction of PAs from a patent database, OAs 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 because there are a number of necessary implementation differences. Section 3.1 details the technical and biological knowledge bases. Thereafter, Section 3.2 explains the prepro-

https://doi.org/10.1017/S0890060415000177 Published online by Cambridge University Press

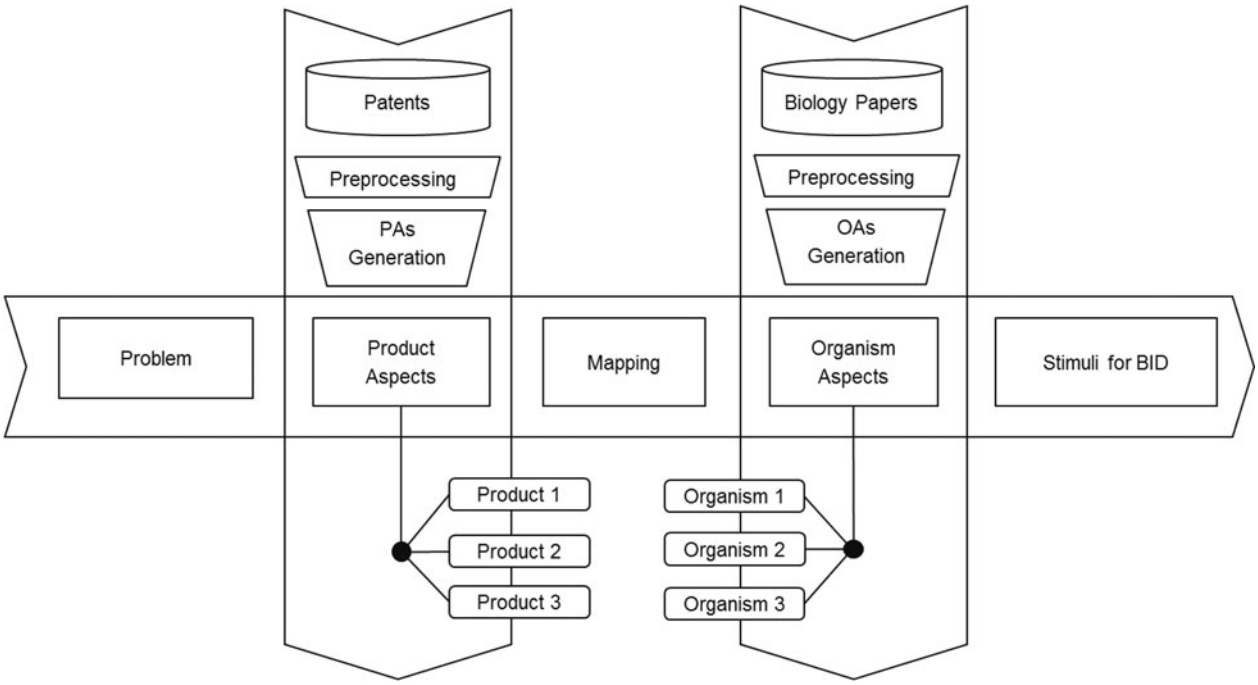

Fig. 1. Overview of the proposed system.

cessing algorithms that transform the biological corpus into a document-term matrix (DTM). Next, in Section 3.3, OAs 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 OAs 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. The full-text descriptions of these patents are used for DTM 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 PA 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 many as possible of the available papers from this journal were taken as the test set, and 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 OA 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. DTM generation

Five preprocessing steps transform each corpus into a 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 (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

https://doi.org/10.1017/S0890060415000177 Published online by Cambridge University Press

Table 2. *Preprocessing example: From corpus document to document vector*

Word	POs Category	Lemma	Filtered by	Document Vector
Study	Noun	Study	Relevance filter	
Investigates	Verb	Investigate	Relevance filter	
Aerodynamic	Adjective	Aerodynamic		1
Gravitational	Adjective	Gravitational		1
Forces	Noun	Force		1
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		1
Acceleration	Noun	Acceleration		1
Diving	Noun	Diving		1

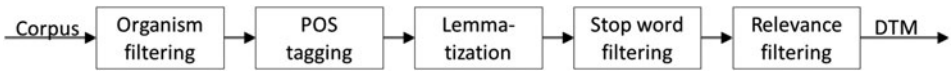

Fig. 2. Preprocessing, from corpus to document-term matrix.

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.

3.2.1. *Organism filtering*

To avoid that OAs 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, D'hondt, et al., 2011) because in both approaches interdocument links are removed to bring out structure relevant for design by analogy. Omitting organism name filtering could result in, for example, an OA 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 OAs would be useless for the envisaged mapping of PAs to OAs.

Organism name detection is performed by LINNAEUS (Gerner et al., 2010), an open-source species name identification system. Its database, containing only names at the species level, is expanded to include all scientific and common organism names of the National Center for Biotechnology Information taxonomy. Because biological strategy documents often contain mentions of ranks higher than the species level, all 26 biological ranks are included (Vandevenne et al., 2015). For the biological corpus of 8011 documents, on average 128 organism mentions are detected per document and 19,782 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 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 and their grammatical function in the sentence, or POS category. After POS 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.

3.2.3. *Lemmatization*

After POS 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. 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

https://doi.org/10.1017/S0890060415000177 Published online by Cambridge University Press

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 DTM, instead of a 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, because they do not represent relevant document content. Examples of stop words are *the*, *and*, *a*, *that*, and *was*. 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 they 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, and objective), all terms representing structure (e.g., wood, body, larva, droplet, oil, wing, and antenna), and all terms that belong to nonrelevant 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 in Table 2.

Because 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 κ (Randolph, 2005) is used because coders are not forced to mark a specific proportion of the terms as relevant. The free-marginal birater κ is calculated to be 0.73. The κ values range from –1 to 1, where –1 indicates perfect disagreement below chance, 0 indicates agreement equal to chance, and 1 indicates perfect agreement above chance. A κ value of 0.7 or above is normally 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 (1932) law 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. Because it can safely be assumed that sporadically occurring terms contribute little to the formation of the principal components (PCs) for OA generation (see Section 3.3), only the terms occurring more than 10 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, result-

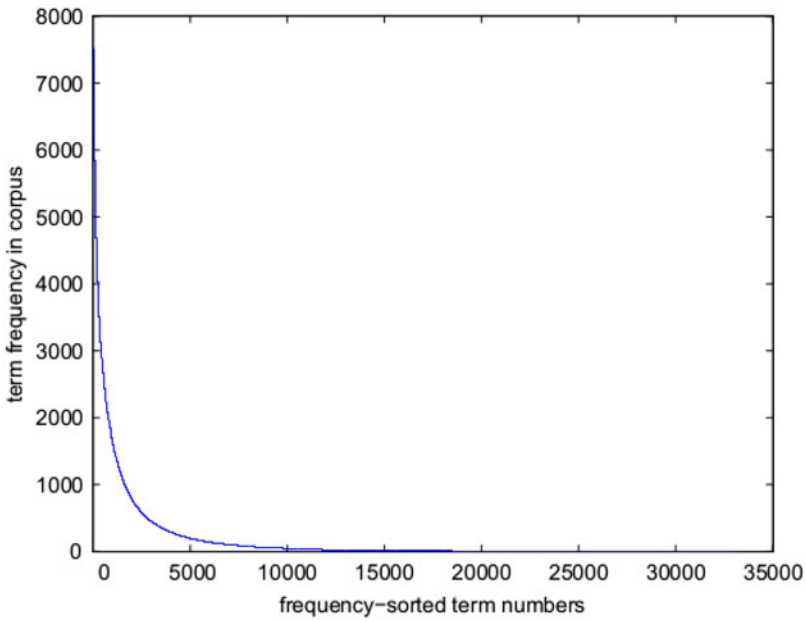

Fig. 3. Illustration of Zipf's law for the corpus dictionary term frequencies.

https://doi.org/10.1017/S0890060415000177 Published online by Cambridge University Press

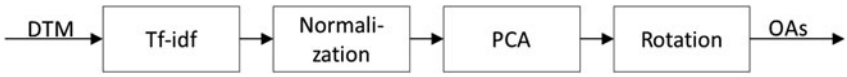

Fig. 4. The organism aspect generation process.

ing 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 because every word occurs exactly once in the sentence.

3.3. PA and OA generation

The OAs generation process is depicted in Figure 4. The DTM 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 DTM is subjected to principal component analysis (PCA; Berry et al., 1995; Skillicorn, 2007). This analysis allows extracting a predefined number of PCs, of which the first PC is the dimension oriented in such a way that it explains the maximum amount of variance in the data set. Each succeeding PC represents as much of the remaining variability as possible, taken into account that all PCs are orthogonal to each other. The authors currently calculate 300 PCs (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.

In a tf-idf weighted DTM, 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 PCs. Furthermore, for the PAnDA tool it has been demonstrated (Verhaegen et al., 2009; Verhaegen, D'hondt, et al., 2011) that Varimax rotation (Kaiser, 1958) facilitates the interpretability of the resulting PCs. After rotation, the PCs are called 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*, and *descend*. Manual labeling 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 OAs. Because such labeling 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 estimating the interrater reliability. Although the raters could use any labeling they seemed fit, most of the OA labels were chosen by picking

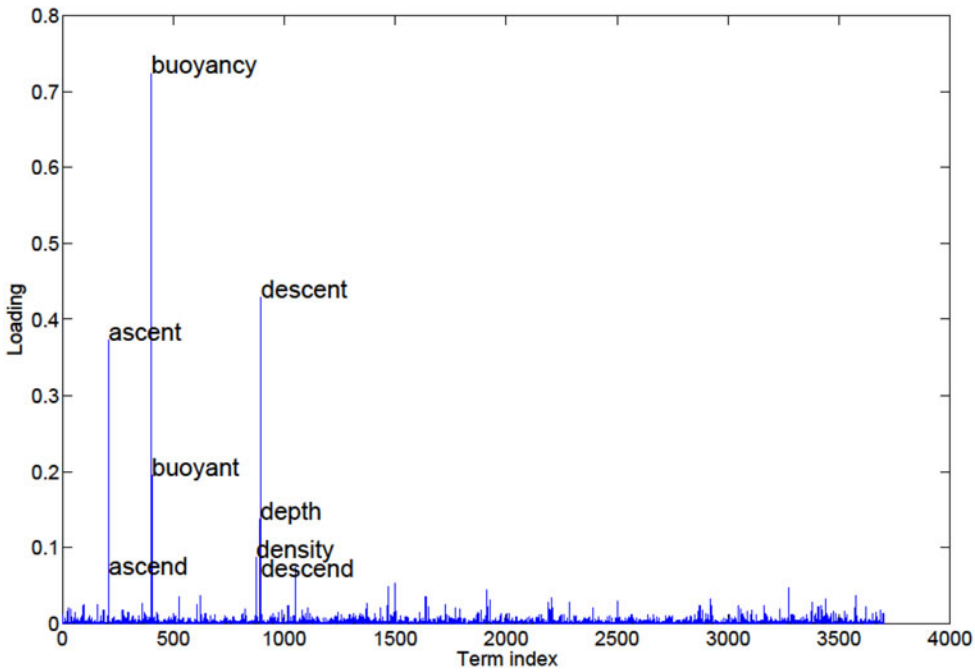

Fig. 5. Term loadings for an example organism aspect.

https://doi.org/10.1017/S0890060415000177 Published online by Cambridge University Press

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

one of the high-loading terms, or a modified form of one of these terms, on the OAs. For this interrater 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 labeled OA dimensions. These two matrices are important information structures that support the functionalities described in Section 4.

3.4. PA and OA mapping

In order to link technical and biological systems for potential knowledge transfer, a similarity measure between the PAs as specified by Verhaegen, D'hondt, et al. (2011) and the above calculated OAs 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, the higher their similarity value will be in the PA-OA matrix. One can distinguish three types of relationships between PAs and OAs:

1. *(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 exists 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 because it is not necessary that all the terms or features in the (near) identical PA and OA vectors are identical.
2. *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 because, 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.
3. *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 PAs and organisms with OAs are similar. The occurrences of products in patents and of organisms in papers need 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, D'hondt, et al., 2011), which consists of 1011 single-word products extracted from the Google product taxonomy (Google Merchant Center, 2014), is improved by a new multiword product identification algorithm, detailed in the paragraphs below. Motivation is twofold: this way much more products are detected (151132), and products can be characterized more precisely. For example, products like *air bag* and *plastic bag*, which would be mapped to the single-word product *bag*, are distinguished by the new multiword product identification algorithm. Such more refined product search improves product characterization results. For example, the most important PAs for *air bag* and *plastic bag* are *inflation/deflation* and *conveying and feeding*, respectively, whereas the most important PA for *bag* is *packaging and sealing*. Without multiword product identification, some of the validation cases presented in Section 5 would not be possible, that is, for the *air conditioning system* and *car body* products.

To identify products, the following POS 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, for example, *valve*, *display*, and *vehicle*. 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, for example, *system*, *device*, and *apparatus*. These are not discarded because the language used in patents often is indirect, such as *communication apparatus* or *packaging system*. A third category contains nouns that have an explanatory function, for example, *temperature*, *pressure*, and *power*; these can explain the words in the previous category, for example, *temperature regulation system*. All other nouns are considered not relevant and placed in the fourth noun category. Adjectives are labeled relevant (e.g., optical, magnetic, and digital) or not relevant (e.g., preferred, first, and improved) 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)

https://doi.org/10.1017/S0890060415000177 Published online by Cambridge University Press

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

resulted in the identification of 151,132 multiword products, which allowed performing validation tests for proof-of-concept purposes (see Section 5). This way, rarely used nouns and adjectives (e.g., *thromboembolism*, *saussurea*, *arrhenius*, *gratuitous*, and *sovereign*) are filtered because the large majority of these infrequently encountered words would be labeled nonrelevant anyway. Noun and adjective categorization is a subjective process; hence, rater agreement is measured. The free-marginal multirater κ (Randolph, 2005) is calculated for noun categorization to four categories and for adjective categorization to two categories. The κ scores for two raters are 0.82 and 0.87, both indicating proof of good interrater agreement (Randolph, 2005).

For multiword product identification, the four noun and two adjective category labels are applied to the retrieved POS sequences as follows:

- If the last word is categorized as nonrelevant (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 labeled as product.
- If the last word is a noun that needs explanation (Noun2) and this explanation is given by either an explaining noun (Noun3), a product noun (Noun1), or a relevant adjective (Adj1), the POS sequence is retained as product.

A final processing step lemmatizes the last words of the retrieved multiword products (e.g., transforming *magnetic strips* to *magnetic strip*) and the frequencies of these lemmatized multiword 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 multiword organism names (e.g., *Lotus japonicus*), as well as common (e.g., gyrfalcon) 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, Verhaegen, et al., 2014).

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. Of course, it is possible that a large fraction of documents for an organism focuses on one specific strategy. One example in the current corpus is the set of 17 documents associated to cuttlefish. Because most of them discuss camouflage, this results in a spike in the radar plot for the OA camouflage. If there is only one document stored for a specific organism, the radar plot typically spikes for one or a few OAs.

4. SEABIRD FUNCTIONALITIES

The algorithms described in the previous section are used to implement the back-end architecture of SEABIRD, as shown

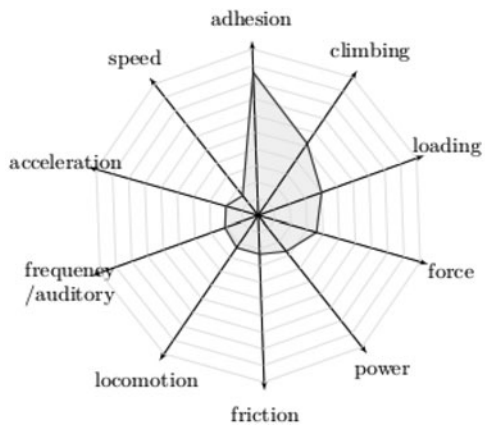

Fig. 6. Radar plot of automatic organism characterization of *geckos*.

https://doi.org/10.1017/S0890060415000177 Published online by Cambridge University Press

in Figure 1. To allow users to interact with the system while moving through the different BID 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 represents a large network containing four types of nodes: products, PAs, OAs, and organisms. Links are formed between these nodes by consulting the characterization and mapping matrices, that is, 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, *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* is characterized with PAs. The selected PA *color perception* maps to OAs from which the OA *color* results in the listed organisms. For the problem-oriented bioinspired design process (Helms et al., 2009), the user navigates from left to right in the network browser.

4.2. Supporting problem formulation: BID process phase 1

Problems are often 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). A direct search allows searching for multiword products. One can search for the exact product or for products starting, ending, or containing a specified character string. The results, which easily contain 10s to 100s of multiword 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*, and so forth. These results also contain product parts; for example, when searching for *air bag*, *air bag cover* and *air bag gas* are also found. Another 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*, and so forth. 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). Because this indirect search also has the potential to generate hundreds of product names, again a frequency-weighted word cloud is used as a visualization method. Product search and selection is an important step in problem formulation, because 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 PAs, 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 realize a specific functionality?* or *How can nature inspire us to realize a specific functionality?* Because there is no product to start from, a direct selection is made from the set of PAs. 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 PA, an ordered list of relevant OAs 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)*.

Table 5. *Representation of the network browser*

Product	Product Aspect	Organism Aspect	Organism
Display	Text & illustration	**Color**	Swallowtails
	Data transmission	Coloration	Strawberry poison frog
	Illumination	Spectrum/reflectance	Steller's jay
	Color perception	Spectral	Monarch butterfly
	Vibrating	Optical	*Cephalotes atratus*

https://doi.org/10.1017/S0890060415000177 Published online by Cambridge University Press

4.3.1. Search for relevant organisms

After selecting an OA in the network browser, the 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 nicobariensisnicobar island frog*, *ribbed mussel*, *bearded seal,* and so forth. 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, because the retrieved organisms are *glowworms*, *Lampyris noctilucacommon*, *dinoflagellates*, *Suberites domuncula*, *fireflies*, *brittle stars*, *Photuris*, and so forth.

4.3.2. Search for relevant strategy documents

For a selected OA, 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. In addition, 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 not only 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 focuses on high-frequency gecko communication. Because 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 analyzed 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,* and so forth. These terms and their inflections (e.g., adhesives, adhesiveness, adhesion, and 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

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, natures 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

Fig. 7. Annotation for an example strategy document for the organism aspect adhesion (Autumn et al., 2006).

https://doi.org/10.1017/S0890060415000177 Published online by Cambridge University Press

that the terms scoring high on the relevant OAs are interesting to focus on during knowledge transfer.

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 phases. 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 bioinspired 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 bioinspired designs are taken, often used as examples for introducing BID, for example, the butterfly-inspired Mirasol display technology (Mirasol, 2009). By taking widely accepted examples of BID as validation cases, the authors aim at maximizing the likelihood of readers being familiar with the presented validation cases and at minimizing doubt whether the bioinspired design was really inspired by nature.

As explained in Section 3.1, SEABIRD currently integrates a relatively small database compared to a 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 OA space with folding-in (Deerwester et al., 1990). Adding biological strategy documents this way does not impede validation because 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 labeling or classification tasks detailed in Section 3 and without the need to regenerate the OAs.

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 coloring of butterflies. According to Vukusic (2006),

> 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.

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 multiword 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*, and so forth. *Display* is taken as the product to characterize, which results in the following ordered PAs: *text and illustration*, *data transmission*, *illumination*, *color perception*, *vibrating*, and so forth. The other high-ranking product synonyms are confirmed to have similar characterizations, mainly differing in the order of the PAs. For the problem at hand, the PA *color perception* is chosen to map to OAs. SEABIRD's PA-to-OA-mapping suggests in decreasing order of importance: *color*, *coloration*, *spectrum and reflectance*, *spectral*, *optical*, *illumination*, *camouflage*, and so forth, and the first OA *color* is selected to explore for relevant organisms. A list of organisms is generated that, 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*, and so forth. Based on SEABIRD's current corpus, 5 of the top 10 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 *color*, the 11th strategy document is titled "Significance of a basal melanin layer to production of noniridescent structural plumage color: Evidence from an amelanotic Stellers jay," discussing structural coloration for birds. For the OA *coloration*, the 9th result is titled "Spectral reflectance and directional properties of structural coloration in bird plumage" and for the OA *spectrum/reflectance* the 2nd, 8th, and 14th result are titled "Glass scales on the wing of the swordtail butterfly *Graphium sarpedon* act as thin film polarizing reflectors," "Blue integumentary structural colors in dragonflies (Odonata) are not produced by incoherent Tyndall scattering," and "Anatomically diverse but-

https://doi.org/10.1017/S0890060415000177 Published online by Cambridge University Press

terfly scales all produce structural colors by coherent scattering," respectively. Each of these titles describes 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 coloring 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). SEABIRD, only relying on the biological papers of one journal, retrieves two strategy documents describing structural coloring 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 multiword products in the patent database containing the word *car*. The most important are *car*, *elevator car*, *electric car*, *railway car*, *motor car*, and so forth. Hence, product search is focused on the part of the car most relevant to aerodynamics, being the *car body*. This product is found with an indirect product search, see Section 4.2, locating products that frequently co-occur with the product *car*. SEABIRD's characterization 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*, and so forth. 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 inspiration,[1] the biological understanding of temperature control in termite mounds had two main components: less dense, warm internal air rises and is exchanged in the chimneys with denser cooler air that moves downward; and 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 a 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 because this functionality does not require an identified organism in the title.

SEABIRD's multiword product search returns a number of candidate products, for example, *air conditioning system*, *air conditioning apparatus*, and *air conditioning device*. The most important PA 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 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. Therefore, again for this case, SEABIRD's search is able to retrieve stimuli that have the potential to trigger a bioinspired 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 bioinspired innovations. The con-

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

https://doi.org/10.1017/S0890060415000177 Published online by Cambridge University Press

Table 6. *Biological strategy documents for air conditioning of buildings, and their rank on the organism aspect* 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

fidence 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 in which 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 (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 interactive tasks required together approximately 40 h to set up SEABIRD. 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 stability should be reevaluated because language itself evolves. If necessary, a small effort can integrate new terms. A similar reasoning holds for the interactive labeling of OAs. By comparing the OA set for a gradually increasing database, a minimal corpus size will be determined where the OA set becomes stable, as previously illustrated for the PA set of the PAnDA tool (Verhaegen, D'hondt, et al., 2011). 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. Because these tasks require a relatively small human effort during SEABIRD's setup, they are not the focus of future automation attempts.

The above interactive tasks during SEABIRD's initial setup require human judgment and are thus subjective and even error prone. Although the confirmed repeatability gives some confidence about the execution of these tasks, it also indicates that opinions for term labeling or classification can differ. The effect of this, however, is limited because of 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 bioideation tool. A further illustration of the resistance of the functioning of the proposed tool to changes in the retained dictionary is that only the terms that occur at least 10 times in the corpus are interactively evaluated; the rest do 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 on finding the right keywords, but on 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*, and *mussels*; and in the patent database, many products that have a functionality related to adhesion or attachment are identified (fas-

https://doi.org/10.1017/S0890060415000177 Published online by Cambridge University Press

tener 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 OAs than the currently arbitrarily chosen sets of 300. Another feature that would benefit SEABIRD is the ability of multiple selection of products, PAs or OAs. This way synonymous products (e.g., computer screen and computer display) could be characterized as one, or multiple PAs could be combined to map to OAs, or multiple OAs 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 recognize these solutions is not yet tested because this task requires solving the second important challenge of effective stimuli representation for supporting knowledge transfer. Identifying valid analogies in natural-language text resulting from search is not a trivial task (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 BID 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 & Markham, 1997). Furthermore, it is likely that instantiating knowledge transfer models (FB, SBF, 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 bioinspired 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, SEABIRD is proposed, a bioideation 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 OAs, two concept sets extracted from, respectively, a patent and biological paper database. Problem formulation is supported by combining an advanced multiword product search with product characterization, resulting in one or more PAs to focus on. A mapping of product to OAs results in an 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. 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 for 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 bioinspired concept generation.

REFERENCES

Altshuller, G.S. (1984). *Creativity as an Exact Science: The Theory of the Solution of Inventive Problems*. New York: Gordon & Breach Science Publishers.

Autumn, K., Dittmore, A., Santos, D., Spenko, M., & Cutkosky, M. (2006). Frictional adhesion: a new angle on gecko attachment. *Journal of Experimental Biology 209(18)*, 3569–3579.

Bajželj, B., Allwood, J.M., & Cullen, J.M. (2013). Designing climate change mitigation plans that add up. *Environmental Science and Technology 47(14)*, 8062–8069.

Bar-Cohen, Y. (2006). Biomimetics—using nature to inspire human innovation. *Bioinspiration & Biomimetics 1(1)*, 1–12.

Bar-Cohen, Y. (2011). *Biomimetics: Nature-Based Innovation*. New York: 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. *Journal of Experimental Biology 208*(Pt. 2), 327–344.

Benyus, J.M. (1997). *Biomimicry: Innovation Inspired by Nature*. New York: 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. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 367(1893)*, 1445–1486.

Bonser, R.H.C. (2006). Patented biologically-inspired technological innovations: a twenty year review. *Journal of Bionic Engineering 3(1)*, 39–41.

Bonser, R.H.C., & Vincent, J.F.V. (2007). Technology trajectories, innovation, and the growth of biomimetics. *Journal of Mechanical Engineering Science 221(10)*, 1177–1180.

Brants, T. (2000). TnT: a statistical part-of-speech tagger. *Proc. 6th Conf. Applied Natural Language Processing, ANLC '00,* pp. 224–331. Stroudsburg, PA: Association for Computational Linguistics.

Chakrabarti, A., Sarkar, P., Leelavathamma, B., & Nataraju, B.S. (2005). A functional representation for aiding biomimetic and artificial inspiration

https://doi.org/10.1017/S0890060415000177 Published online by Cambridge University Press

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. (2012). Understanding analogical reasoning in biomimetic design: an inductive approach. *Proc. Design Computing and Cognition Conf., DCC'12* (Gero, J.S., Ed.). Berlin: Springer.
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*. Advance online publication. doi:10.1115/1.4027494
Chiu, I., & Shu, L.H. (2005). Bridging cross-domain terminology for biomimetic design. *Proc. ASME 2005 Int. Design Engineering Technical Conf.*, Paper No. DETC2005-84908, Long Beach, CA, September 24–28.
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., McAdams, D.A., & Stone, R.B., Eds.), pp. 17–27. London: Springer–Verlag.
Fall, C.J., Törcsvári, 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. (1989). 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. *Journal of Mechanical Engineering Science 223(12)*, 2899–2918.
Gentner, D., & 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.* Accessed at https://support.google.com/merchants/answer/160081?hl=en on May 14, 2008.
Helms, M., Vattam, S.S., & Goel, A.K. (2009). Biologically inspired design: process and products. *Design Studies 30(5)*, 606–622.
Intergovernmental Panel on Climate Change. (2007). *IPCC Fourth Assessment Report Working Group I Report: The Physical Science Basis*. New York: Cambridge University Press.
Kaiser, H.F. (1958). The Varimax criterion for analytic rotation in factor analysis. *Psychometrika 23(3)*, 187–200.
Kaiser, M.K., Farzaneh, H.H., & Lindemann, U. (2012). An approach to support searching for biomimetic solutions based on system characteristics and its environmental interactions. *Proc. Design 2012*, pp. 969–978, Cavtat-Dubrovnik, Croatia, September 24–28.
Kaiser, M.K., Farzaneh, H.H., & Lindemann, U. (2014). BIOscrabble—the role of different types of search terms when searching for biological inspiration in biological research articles. *Proc. Design 2014*, pp. 241–250, Cavtat-Dubrovnik, Croatia, May 19–22.
Landauer, T.K., & 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. *Proc. 4th ACM Conf. Digital Libraries*, pp. 179–187. New York: ACM Press.
Lenau, T., Dentel, A., Ingvarsdóttir, Þ., & Guðlaugsson, T. (2010). Engineering design of an adaptive leg prosthesis using biological principles. *Proc. Design 2010*, pp. 331–340, Cavtat-Dubrovnik, Croatia, May 17–20.
Lepora, N.F., Verschure, P., & Prescott, T.J. (2013). The state of the art in biomimetics. *Bioinspiration & Biomimetics 8(1)*, 013001.
Linnaeus, C. (1767). *Systema Naturae*. Amsterdam: Author.
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]. Accessed 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. University of Texas at Austin, Department of Mechanical Engineering.
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. *Proc. 39th ASEE/IEEE Frontiers in Education Conf.*, pp. 1–5, San Antonio, TX, October 18–21.
Purves, W.K., Sadava, D., Orians, G.H., & Heller, H.C. (2001). *Life: The Science of Biology*, 6th ed., Sunderland, MA: Sinauer Associates.
Randolph, J.J. (2005). Free-marginal multirater kappa: an alternative to Fleiss' fixed-marginal multirater kappa. *Proc. Joensuu University Learning and Instruction Symp.*, Joensuu, Finland.
Salton, G., & Buckley, C. (1988) Term-weighting approaches in automated text retrieval. *Information Processing & Management 24(5)*, 513–523.
Salton, G., Wong, A., & Yang, C.S. (1975). A vector space model for automatic indexing. *Communications of the ACM 18(11)*, 613–620.
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.
Shu, L.H. (2010). A natural-language approach to biomimetic design. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing 24(4)*, 507–519.
Skillicorn, D. (2007). *Understanding Complex Datasets: Data Mining with Matrix Decompositions*. New York: Chapman & Hall.
Srinivasan, V., Chakrabarti, A., & Lindemann, U. (2012). Towards an ontology of engineering design using SAPPhIRE model. *Proc. CIRP Design 2012*, pp. 17–26. London: Springer.
Turner, S., & Soar, R. (2008). Beyond biomimicry: what termites can tell us about realizing the living building. *Proc. 1st Int. Conf. Industrialized, Intelligent Construction (13CON)*, pp. 221–237, Loughborough, May 14–16.
Vandevenne, D., Verhaegen, P.-A., & Duflou, J.R. (2014). *Methods and algorithms for systematic biologically-inspired design.* PhD Thesis. KU Leuven.
Vandevenne, D., Kellens, K., Banck, M., Clijsters, H., Verhaegen, P.-A., & Duflou, J.R. (2012). A framework for assessing the sustainability of biologically-inspired products. *Proc. Innovation for Sustainable Production*, Bruges, Belgium, May 6–9.
Vandevenne, D., Verhaegen, P.-A., Dewulf, S., & Duflou, J.R. (2011). A scalable approach for the integration of large knowledge repositories in the biologically-inspired design process. *Proc. 18th Int. Conf. Engineering Design (ICED11) 6*, pp. 210–219, Copenhagen, Denmark, August 15–19.
Vandevenne, D., Verhaegen, P.-A., Dewulf, S., & Duflou, J.R. (2014). A scalable approach for ideation in biologically inspired design. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*. Advance online publication. doi:10.1017/S0890060414000122
Vandevenne, D., Verhaegen, P.-A., & Duflou, J.R. (2015). Mention and focus organism detection and their applications for scalable systematic bio-ideation tools. *Journal of Mechanical Design*, 136, Article 111104.
Vattam, S.S., Helms, M.E., & Goel, A.K. (2010). A content account of creative analogies in biologically inspired design. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing 24(4)*, 467–481.
Vattam, S.S., Wiltgen, B., Helms, M., Goel, A.K., & Yen, J. (2010). DANE: fostering creativity in and through biologically inspired design. *Proc. Int. Conf. Design Creativity*, pp. 115–122, Kobe, Japan, November.
Vattam, S.S., & Goel, A.K. (2011). Semantically annotating research articles for interdisciplinary design. *K-CAP'11*, pp. 165–166. New York: ACM.
Verhaegen, P.-A., D'hondt, J., Vertommen, J., Dewulf, S., & Duflou, J.R. (2009). Quantifying and formalizing product aspects through patent mining. *Proc. ETRIA TRIZ Future 2009*, Timisoara, Romania.
Verhaegen, P.-A., D'hondt, J., Vandevenne, D., Dewulf, S., & Duflou, J.R. (2011). Identifying candidates for design-by-analogy. *Computers in Industry 62(4)*, 446–459.

Verhaegen, P.-A., & Duflou, J.R. (2013). *Methods and algorithms for systematic innovation.* PhD Thesis. KU Leuven.
Verhaegen, P., Peeters, J., Vandevenne, D., Dewulf, S., & Duflou, J.R. (2011). Effectiveness of the PAnDA ideation tool. *Procedia Engineering 9*, 63–76.
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*. Cambridge, MA: Harvard University Press.

Dennis Vandevenne is a Postdoctoral Researcher on methods and algorithms for BID at the Centre for Industrial Management and a Researcher at Katholieke Universiteit Leuve (KU Leuven). He obtained three information and communications technology (ICT)-related master's degrees: electronics–ICT, artificial intelligence–engineering and computer science, and industrial management–ICT. Dennis previously performed research on identity management and biometrics in the Department of Computer Security and Industrial Cryptography.

Paul-Armand Verhaegen is a Postdoctoral Researcher in the domain of systematic innovation, product development, and data mining at KU Leuven. He holds a master's degree in applied science and engineering, with a specialization in electrotechnical and computer science. Paul-Armand is a graduate in complementary studies in business administration; he attained an MBA from Vrije Universiteit Brussel and a PhD in engineering from KU Leuven. Dr. Verhaegen founded Stocks, a company specializing in printer supplies. He worked as an external consultant for 3E on green energy certificates; as a consultant at the Bureau van Dijk Management Consultants; and as a Research Assistant at Vrije Universiteit Brussel, Erasmushogeschool Brussel, and KU Leuven.

Simon Dewulf is Founder of CREAX. CREAX developed Creation Suite (http://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 GlaxoSmithKline to provide a direct access to out-of-domain knowledge for technology transfer and open innovation opportunities.

Joost R. Duflou is a Professor in the Department of Mechanical Engineering at KU Leuven. He holds master degrees in architectural and electromechanical engineering and a PhD in engineering from KU Leuven. 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. Dr. Duflou is a member of CIRP and has published over 200 international publications.