DETC2012-71296

EXPLORING THE COLLECTIVE CATEGORIZATION OF BIOLOGICAL INFORMATION FOR BIOMIMETIC DESIGN

Jungik (Jay) Son
Dept. of Mechanical and Industrial Engineering
University of Toronto
Toronto, ON, Canada

Christian Raulf[+]
Department 5, Nature and Technics
University of Applied Sciences Bremen
Bremen, Germany
[+] Paper describes work performed at University of Toronto

Hyunmin Cheong
Dept. of Mechanical and Industrial Engineering
University of Toronto
Toronto, ON, Canada

L. H. Shu*
Dept. of Mechanical and Industrial Engineering
University of Toronto
Toronto, ON, Canada
Corresponding author email: shu@mie.utoronto.ca

ABSTRACT

Categorizing biological information can be subjective and ambiguous, which poses challenges for indexing potentially useful biological information for design. Therefore, we explored collective categorization to study the categorization task. After gathering 163 examples of biological transformation, we asked four participants to independently categorize the examples using self-selected approaches. A computational algorithm was used to quantify the relatedness between the groups that each participant created. The results confirmed that participants had different perspectives in interpreting and categorizing biological information. However, the collective categorization method could reveal meaningful semantics in biological information such as hierarchical, synonymous, or causal relations. The relations discovered could lead to developing formal representations or learning unique patterns in biological phenomena.

1. INTRODUCTION

One challenge in biomimetic design is developing generalized methodologies to identify and apply relevant biological analogies. To identify analogies, our research group has developed techniques to extract relevant information from sources in natural-language format, e.g., text, papers, etc. (Shu, 2010). This approach allows designers and researchers to access a large amount of readily available biological information, rather than require the creation of databases.

Another approach is to create formal representations to support cataloging biological information in databases (Chakrabarti et al. 2005, Goel et al. 2009, Nagel et al. 2010). The formal representations could help designers comprehend the biological information, and also allow machines to reason with the information, e.g., to support design-by-analogy.

The two approaches could well support each other. The natural-language approach could identify useful information to be entered into databases. The formal representations developed could help researchers understand the type of information to be captured from natural-language text.

However, categorizing biological information for entry into a database is challenging and subject to researcher bias. In addition, modeling complex biological systems with representations developed for engineering applications can be ambiguous. Evaluating and validating such models present challenges as well.

Therefore, we investigate a new categorization method to address the challenges highlighted above. Our approach combines categorization results from multiple researchers with a computational algorithm to examine relatedness in their categorizations. We describe our approach and report benefits and challenges observed during the categorization process.

The biological information collected and categorized focused on transformation that occurs in nature. Transformation in mechanical products provides unique benefits that products otherwise cannot, e.g., an umbrella that

folds to enhance portability, an aircraft that changes its wing shape to adapt to different modes of flight. A number of design researchers (Singh et al. 2009, Camburn et al. 2010, Kuhr et al. 2010) have examined existing transformable products and developed techniques to assist in designing transformable products. In particular, Singh et al. (2009) examined transformation in biological systems and were able to draw analogies between transformation in nature and engineering products. We focused on investigating transformation in biological phenomena, with the eventual goal of identifying interesting patterns of biological transformation that may be applicable to engineering design.

2. BACKGROUND

2.1. Study of transformation in design

Singh et al. (2009) examined existing products, patents, and biological phenomena to extract three transformation principles and twenty transformation facilitators. The three transformation principles, or "generalized directive[s] to bring about a certain type of mechanical transformation," include: 1) expand / collapse, 2) expose / cover, and 3) fuse / divide. Twenty transformation facilitators, e.g., flip, fold, material flexibility, etc., describe mechanisms or characteristics that enable transformation. Singh et al. also identified analogous biological phenomena for each principle and facilitator.

Camburn et al. (2010) and Kuhr et al. (2010) developed techniques and tools to support the design of transformable products. Camburn et al. (2010) derived a set of indicators to identify when implementing transformation in products may be beneficial. Kuhr et al. (2010) developed a method to help designers examine the functional states of single-function devices and apply relevant transformation facilitators.

2.2. Methods to support biomimetic design

2.2.1. Search natural-language text

Shu (2010) summarized the natural-language approach, which involves searching biological information in natural-language format for relevant analogies. The approach enables designers to identify analogies that are not limited to those suggested by biologists or indexed by design researchers. However, challenges of the approach include: 1) lexical differences between engineering and biology and 2) a potentially large number of search results to manage.

Chiu and Shu (2007) and Cheong et al. (2011) developed a process for identifying biologically meaningful keywords to address the first challenge. Such keywords may locate more relevant analogies in biological text than the corresponding engineering keywords. For the second challenge, Ke et al. (2010) applied natural-language processing to categorize search results. However, challenges also exist in identifying and transferring analogies from descriptions of biological phenomena to engineering solutions (Mak and Shu 2004, 2008; Cheong et al. 2010).

2.2.2. Model and index biological systems

A number of researchers have developed formal representations to model and index biological systems, which can help designers understand complex biological information for design-by-analogy. Chakrabarti et al. (2005) developed SAPPhIRE constructs to represent causality in natural and artificial systems and created a database of these systems. Goel et al. (2009) used the structure-behavior-function (SBF) model to represent biological phenomena and created a library, IDEAL, that contains the SBF models. Nagel et al. (2010) used the functional basis to model biological systems and index them in a design repository.

These formal representations use terminologies or frameworks that are abstract or engineering-oriented, facilitating the transfer and application of biological knowledge to design problems. However, a primary limitation of the modeling approach is that it requires resources and expertise to construct and categorize information for the models, a process that may introduce researcher bias.

2.2.3. Generalize patterns of biological solutions

Instead of modeling individual biological systems, one could generalize patterns in strategies used by biological systems. Altshuller (1984) identified inventive principles based on patterns of innovation found in patents, resulting in the TRIZ matrix, which provides innovative principles to solve contradicting objectives in design problems. Vincent et al. (2006) developed BioTRIZ, which rearranged the original TRIZ matrix based on how contradicting objectives are solved in over 500 biological phenomena.

Vandevenne et al. (2011) proposed a scalable approach to identify the patterns of term occurrences in natural-language text. The approach could bridge the gap between directly searching natural-language text for relevant information and identifying general patterns of biological solutions. Cheong and Shu (2012) discuss the automatic extraction of causally related functions from natural-language text, which can also contribute towards identifying patterns of biological strategies.

2.3. Information categorization techniques

2.3.1. Categorization for biomimetic design methods

Each method developed to support biomimetic design uses a different approach to categorize biological information.

Initially, our natural-language approach allowed designers to use and categorize retrieved information to suit them. More recently, we use syntactic or semantic information (Ke et al. 2010, Cheong and Shu 2012) to categorize retrieved information towards identifying the most relevant information.

Vandevenne et al. (2011) propose a scalable approach to categorize biological information, mainly based on the lexical relationships between documents containing the biological information. This approach can be highly automated and is therefore well suited to categorizing a large amount of information. However, it may overlook the semantics and context of the biological information.

The modeling approaches require researchers to manually enter information, as well as develop formal representations to capture information. As such, axiomatic or empirical validation of the representation schemes is essential.

The above challenges motivated us to develop a categorization method that is compatible with both manual and computational categorization.

2.3.2. Formal representations vs. tagging

The main benefit of formal representation is that information can be captured in a consistent format that facilitates machine reasoning. Goel (1997) and Umeda and Tomiyama (1997) argue that formal representations enable artificial intelligence to support innovative design, including analogical design. While increased formality minimizes ambiguity and helps machines understand information, it often results in loss of flexibility in capturing information (Baclawski and Niu 2006). This can be particularly problematic while modeling complex biological information.

Tags, or metadata, are non-hierarchical keywords that are assigned to describe particular data. Tagging is highly flexible because taggers are free to label information in any way. One well-known use of tagging is on web blogs where the authors label free-form tags on their entries, which help authors organize, and others find, tagged entries. However, since tags are freely chosen, the main disadvantage is semantic ambiguity.

A folksonomy is a collaborative taxonomy that emerges from tags created and shared by multiple people. Halpin et al. (2007) and Robu et al. (2009) demonstrated that collaborative taxonomy can self-organize and converge toward a more controlled taxonomy.

For our work, we aim to explore the collective categorization that can emerge from categories and tags used by multiple people. The next section describes our methods.

3. METHODS

Figure 1 gives an overview of our methods. First, we collected 163 examples of biological transformation while defining the characteristics and scope of the information collected. Four participants then independently categorized the examples and tagged each group with keywords. An algorithm was developed to analyze the tagging data.

3.1. Defining bio-transformation

While collecting examples of natural transformation in the initial exploration phase, we discovered that we must define our scope. Singh et al. (2009) defined transformation as "the act of changing state in order to facilitate new, or enhance an existing functionality." Because this definition is intended to achieve wide coverage in the engineering domain, it does not capture specific details required to distinguish subtle differences between various physical changes in biology. Therefore, we developed a set of requirements to define the scope of transformation examples collected for this research:

Requirement 1: possession of at least two stable states
Requirement 2: change of function across different states

Some examples of biological phenomena we excluded include: 1) evolutionary changes, 2) growth patterns, and 3) processes that are not completely understood. First, despite potentially inspiring useful ideas for design, evolutionary changes in organisms have a very long timeline, and may not be as relevant for designing artificial transformers that must readily alternate between functions. We also disregarded examples of metamorphosis or plant growth that do not involve distinct states. Finally, we excluded examples where we did not have a complete understanding of the exact mechanisms involved, e.g., DNA replication.

This process revealed that even identifying relevant characteristics and limitations for the scope of biological information to include is a relatively subjective task. We suspect that this can present significant challenges in the modeling approaches to biomimetic design.
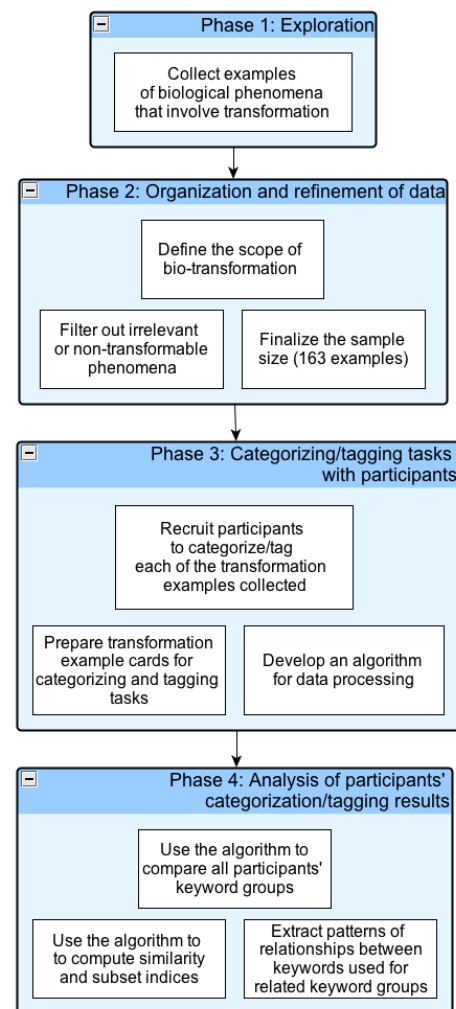
Figure 1: Overview of methods.

3.2. Methods used to collect examples

The following search methods were used to collect examples of relevant biological transformation:

1. Recall personal knowledge and search for confirming evidence (~40% of examples)
   - Examples: A salamander amputating its tail, Venus flytrap closing its leaves, a chameleon changing its color.

2. Use keywords in public search engines (~25% of examples)
   - Example keywords used:
     - Nouns: adaptation, adaptive behavior, autonomy, self-amputation, camouflage.
     - Verbs: fold, shape, fuse, divide, change.

3. Explore websites containing collections of biological phenomena (~25% of the examples)
   - Examples of sources accessed: The Encyclopedia of Earth (EOE) website, video collections on BBC Nature, e.g., movie clips in the "Animal and Plant Adaptations and Behaviour" section.

4. Miscellaneous (~10% of the examples)
   - Discuss biology with subject matter experts.
   - Review literature, e.g., Hirnschall (1975).
   - Survey museum exhibits, e.g., the natural history section at the Royal Ontario Museum in Toronto.

Table 1 shows examples of resources we used. Once relevant information on transformation was found, other scientific journals or websites were researched for more complete descriptions of the information, i.e., the underlying mechanism of the corresponding transformation.

Table 1: Primary sources used to gather examples of biological transformation.

Source type	Source
Online encyclopedias	– Animal Diversity Web (http://animaldiversity.ummz.umich.edu) – BBC Nature (www.bbc.co.uk/nature) – Encyclopedia of Earth (www.eoearth.org) – Encyclopedia of Life (http://eol.org) – Wikipedia (http://en.wikipedia.org)
Biomimicry Institute portal	– Ask Nature (www.asknature.org)
Scientific journals	– Nature (www.nature.com) – Science (www.sciencemag.org)

The initial search effort identified 80 examples of natural transformation, but this effort focused on popular sources of information. To include other relevant examples, we reviewed a reference text for an entry-level university biology course, *Life: The Science of Biology*. Searching this source, as well as revisiting other primary sources in the initial phase, provided 83 additional examples of transformation. Overall, 163 relevant examples were collected for this study.

One challenge encountered was the time-consuming nature of the search process. In many cases, the actual mechanisms of transformation were not fully described, requiring further research to determine if a phenomenon was actually relevant to the scope of this study. The presence of certain keywords, e.g., fold, change color, etc., in the description often distinguished whether the biological phenomenon involved relevant transformation behavior.

We recognize that bias could skew the examples collected. For instance, because many sources that we studied described phenomena at the organ to organism level (e.g., behaviors of animals or plants), potentially useful transformation examples at other levels (e.g., the molecular to cellular levels) may have been neglected.

3.3. Initial categorization of examples

At first, two of the authors looked for patterns in the examples of biological transformation, and found 20 transformation principles that organisms use to achieve 21 objectives. These principles and objectives were determined after thorough discussion and mutual agreement between the two. These discussions were beneficial in enriching the understanding of the transformation examples. When disagreement could not be resolved, more information was sought until both agreed on how to categorize the example.

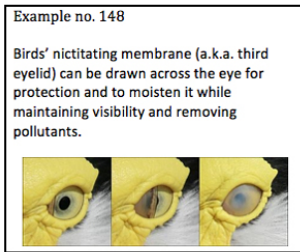This process highlighted an intrinsic limitation in manual categorization, that of subjectivity.

3.4. Tagging examples with multiple participants

Next, four additional people independently categorized the transformation examples and tagged them with keywords. Two were authors of this paper, not involved in the information collection process, while the other two were colleagues: graduate students researching design theory and methodology. We hypothesized that incorporating multiple perspectives in the categorization process would reveal interesting patterns of transformation. We also wanted to explore the collective taxonomies or categories that could emerge from this process.
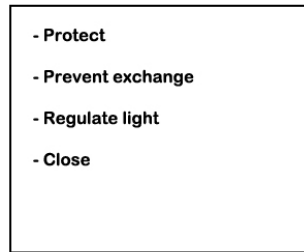
The examples collected were prepared as individual cards, each containing a description and a picture of a specific biological transformation. Each participant received all 163 cards. Figure 2 shows an example card.

Participants were instructed to 1) group the cards based on any attributes they perceived as important from the examples and 2) tag each card with keywords to describe important attributes that they noted. Participants were free to choose whichever categories, keywords, or order of performing the two tasks that they preferred. They could also use multiple keywords to describe each card. Participants were reminded that the eventual goal of the study was to extract patterns of biological transformation, so that they would not categorize the examples based on taxonomical classifications, e.g., genus or species. The goal of this categorization/tagging task was to capture the participants' individual perspectives on patterns in the given set of transformation examples.

Transformation example card - front side

Example no. 148

Birds' nictitating membrane (a.k.a. third eyelid) can be drawn across the eye for protection and to moisten it while maintaining visibility and removing pollutants.

Back side with Participant A's keywords

- Protect

- Prevent exchange

- Regulate light

- Close

Figure 2: An example card and keywords tagged by Participant A.

The resulting data from each participant consisted of a set of examples described by specific keywords. For example, Participant A used the keyword, "absorb", to describe Examples 24, 69, and 133. We interpreted this as participant A creating a group (or category), labeled "absorb". Because multiple keywords could be assigned to each example, groups were not mutually exclusive. After the categorization/tagging task, participants were debriefed and strategies used to carry out the task were discussed.

3.5. Developing algorithm to analyze categorization results

The lead author preprocessed the categorization results to ensure that keywords describing the same attribute were used in a consistent grammatical format, e.g., change "absorption" to "absorb". The resulting data set contained pairs of example identification numbers and keywords, e.g., "24-absorb".

Han and Chen (2007) and Robu et al. (2009) investigated models to analyze collaborated tagging from a large number of people in information systems. We developed a simpler algorithm to enable accurate analysis of a relatively small data set, rather than approximate trends in a large amount of data.
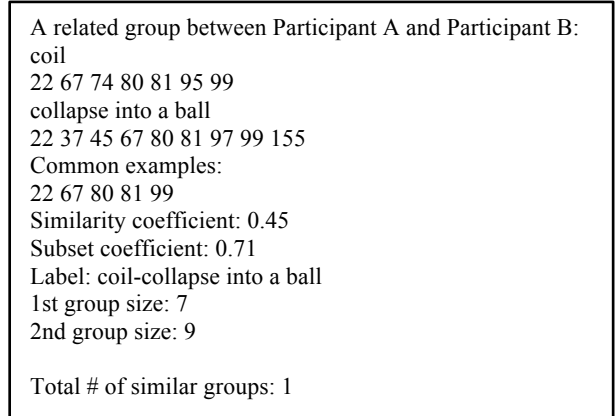
The algorithm essentially identified the relatedness of groups between multiple participants. Two different measures were used to determine the relatedness: 1) subset index, the ratio of the size of intersection between two groups to the size of the smaller group and 2) similarity index, the ratio of the size of intersection between two or more groups to the size of their union. The two indices are expressed as:

$$\text{SU (subset index)} = \frac{n(group1 \cap group2)}{n(groupX)}$$

$$\text{SI (similarity index)} = \frac{n(group1 \cap group2 \cap ...groupN)}{n(group\,1 \cup group\,2 \cup ...groupN)}$$

n( ) = number of examples in the set
groupX = the smaller group of group1 and group2
N = total number of groups being compared

Figure 3 shows an example where Participant A's group has 7 transformation examples, Participant B's group has 9 transformation examples, and the two groups share 5 examples. The subset index would be 5/7 = 0.71, and the similarity index would be 5/(7+9-5) = 0.45.

A related group between Participant A and Participant B:
coil
22 67 74 80 81 95 99
collapse into a ball
22 37 45 67 80 81 97 99 155
Common examples:
22 67 80 81 99
Similarity coefficient: 0.45
Subset coefficient: 0.71
Label: coil-collapse into a ball
1st group size: 7
2nd group size: 9

Total # of similar groups: 1

Figure 3: An example output of the algorithm.

The subset index could potentially identify a hierarchical relationship between concepts describing groups. The similarity index, on the other hand, could identify synonymous concepts used between groups. These two relationships are fundamental to building taxonomies or hierarchies used in information systems (Baclawski and Niu 2006).

The subset index is calculated only between two participants, while the similarity index can be calculated among multiple participants. The algorithm currently calculates the indices for every possible combination of groups among all participants.

The relatedness between groups is based on the common examples contained in the groups, not the keywords used to describe the groups. In other words, when two participants formed similar groups based on a similar attribute but labeled them with different keywords, e.g., "coil" vs. "roll", the algorithm could still determine that those two groups were similar. Thus, the algorithm can identify patterns from multiple participant tags despite different lexicons used to tag.

4. RESULTS

This section reports observations from the participant tagging results, participant interviews, and algorithm output.

4.1. Multiple categorization approaches used by participants

The data collected from participants showed large differences in the number of keywords and categories used between participants. Figure 4 shows the total number of groups and the average number of examples in each group created by the participants. Figure 5 shows the frequency of group sizes each participant created. The following characteristics of each participant's categorization approach were observed:

- Participant A created mostly small groups.
- Participant B created many small groups and a few large groups.
- Participant C had many medium-sized groups.
- Participant D had mostly medium-sized and large groups.

5   Copyright © 2012 by ASME

Figure 4: Total number of groups and average number of examples in groups created by each participant.

Figure 5: Group size frequency of participants.

Participants A and C had few large groups because they did not create super-categories, i.e., categories within categories. In contrast, Participants B and D created small categories within larger categories to form hierarchies.

These observations were consistent with the debrief results, as summarized in Table 2, showing the various strategies that participants used to carry out the categorization/tagging.

As expected, participants used different methods to carry out the categorization/tagging task. For example, Participant A used only the keyword tagging method throughout the sorting task, while Participants C and D mainly used the categorization method. Participant B simultaneously used the tagging and categorization methods.

Table 2: Summary of each participant's categorization/tagging methods and attributes of focus.

Partic.	Primary method	Attributes of focus	Additional note
A	Tagging with keywords	Mechanisms and goals (equal priority)	Did not create super-categories
B	Categorization and tagging with keywords	Mechanisms (priority) and then goals	Combined some categories into super-categories
C	Categorization	Any attribute	Categorized an example under multiple categories; no super-categories
D	Categorization	Goals only	Created super-categories

The participants also focused on different attributes to categorize or tag the examples. For instance, Participant B grouped the examples based on mechanisms of transformation, and then tagged each group with keywords that describe corresponding functional goals achieved by the mechanisms. In contrast, Participant D strictly categorized the examples based on the goals achieved. Participant C's strategy was notably different; this participant created categories based on any attribute that could be associated from the examples.
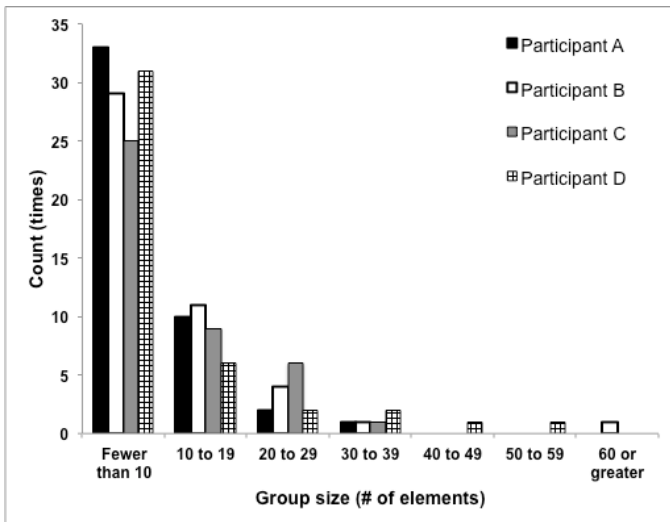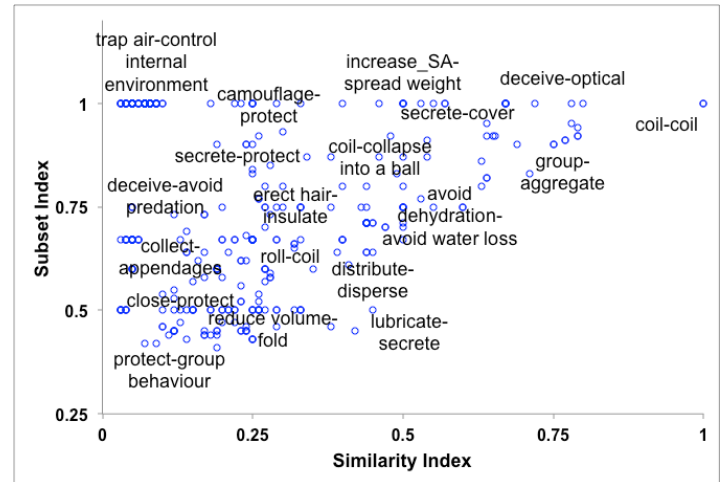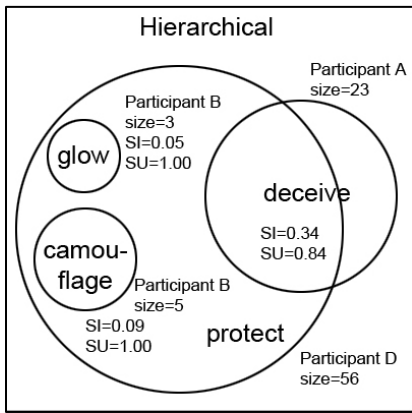
4.2. Relationships identified from categorization results

The algorithm described in Section 3.5 quantified the relatedness between the groups participants generated. In addition to the hierarchical and synonymous relationships identified by the subset and similarity indices, we observed that causal relationships could be inferred from the pairs of related groups identified. The characteristics of each relationship discovered from the computational output are summarized below. Figure 6 depicts these relationships.

- **Hierarchical**: the pair has a high subset index, but a low similarity index. Often, the pair consists of one large-sized group and one small-sized group.
  Ex: "roll" (n=3) and "cover" (n=12)
      subset index = 1.00
      similarity index = 0.25
- **Synonymous:** the pair has both high similarity and high subset indices. This also suggests that the sizes of the two groups are similar.
  Ex: "secrete" (n=26) and "liquid coating" (n=24)
      subset index = 0.92
      similarity index = 0.79
- **Causal:** the pair usually has a high subset index and usually a low-medium similarity index.
  Ex: "deceive" (n=23) and "protect" (n=56)
      subset index = 0.87
      similarity index = 0.34

Hierarchical and synonymous relationships could be determined purely based on the indices, suggesting that these relationships could be automatically identified. However, causal relationships did not have clear-cut criteria and require interpretation of the semantics of the keywords used.

Causal relationships can be identified when participants choose to describe different attributes of the same group. For instance, one participant may use the keyword, "cover", to describe the common mechanism used in a particular group, while another participant may use the keyword, "protect", to describe the common goal of the same group.

Figure 6: Graphical representation of the relationships discovered between participant-created groups.

Figure 7 shows all group pairs between Participants A and B plotted over the corresponding subset and similarity indices. The pairs with high similarity indices near the right end of the plot, e.g., "coil-coil", "group-aggregate", "deceive-optical", etc., exhibit synonymous relationships. The pairs with medium to high subset indices but low to medium similarity indices exhibit either hierarchical, e.g., "roll-coil", or causal relationships, e.g., "trap air-control internal environment", "secrete-protect", "deceive-avoid predation".

Figure 7: Group pairs plotted over the subset and similarity indices. Data points are selectively labeled with keywords used by the participants.

4.3. Examples of relationships identified

Table 3 lists pairs of related groups that scored relatively high subset or similarity indices and exhibit either causal or hierarchical relationships. This type of information could be used to identify important patterns in biological transformation and help develop appropriate relations or categories for formal representations.
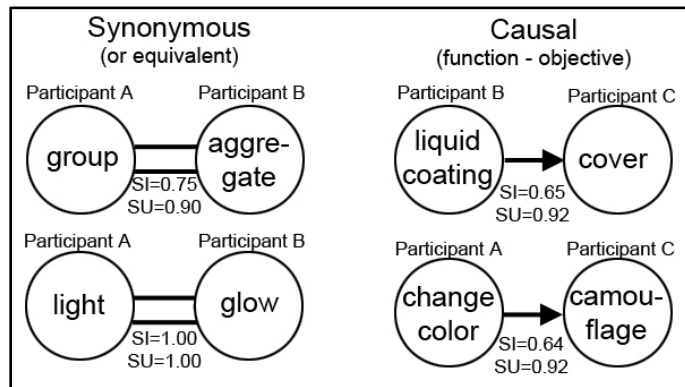
Table 3: Pairs of related groups between participants that exhibit causal or hierarchical relationship.

Related groups	Subset Index	Similar. Index	Relationship
reverse body part-enhance mobility	1.00	0.80	Causal
secrete-prevent	0.92	0.79	Causal
secrete-cover	0.95	0.64	Causal
change color-camouflage	0.92	0.63	Causal/ hierarchical
cover-insulate	0.80	0.40	Causal
deceive-protect	0.87	0.34	Causal
detach-separate	1.00	0.33	Hierarchical
distribute seed-distribute	1.00	0.30	Hierarchical
trap air-insulate	1.00	0.27	Causal
coat-cover	1.00	0.25	Hierarchical
change color-reduce heat absorption	1.00	0.22	Causal

These results, however, are specific to this current study. The categories that emerge from the interpretation of given examples will always be dependent on the examples collected, which is one limitation of such bottom-up category development.

5. DISCUSSION

The categorization process identified challenges in sorting biological information, and how collective categorization and computational tools could help overcome those challenges. We believe that the collective categorization approach could be further automatized to discover patterns from a large amount of biological information.

Copyright © 2012 by ASME

5.1. Challenges in categorizing biological information

As suspected, we observed differences in how each participant categorized examples and identified relevant attributes of the examples. These differences led the algorithm to discover fewer significant relationships than expected. The modeling frameworks developed for biomimetic design are intended to achieve consistency in knowledge representation and categories created. However, the consistency achieved between multiple people modeling/categorizing biological information is often not discussed in the evaluation of those models. More importantly, the models are usually created by a small group of researchers; consequently, bias could exist in the representation schemes used, e.g., attributes or relations defined to describe the biological information modeled.

Another significant challenge observed was that biological phenomena often involve ambiguous mechanisms or goals. Even with the freedom to categorize/tag the examples as they please, participants reported difficulty with grouping some examples and tagging them with appropriate keywords. If a controlled vocabulary and categories were provided in advance, they may not have been flexible enough to capture some biological information. This is a particularly relevant concern because most modeling frameworks used for biomimetic design were initially developed to represent engineering systems. Vincent and Mann (2002) reported that patterns of some biological solutions could not be mapped to an engineering problem-solving methodology such as TRIZ.

The collective categorization method also has its limitations. As discussed earlier, Participant C in this study took a significantly different approach to categorize the examples. It is difficult to assess whether his categorization results would be useful for forming a collaborative taxonomy, solely because his results were so dissimilar to the majority. In addition, a biologist or other nonengineer is likely to create a significantly different categorization scheme. Future studies of the collective categorization method should ensure that these unique inputs are not "smoothed out" from the results of the majority, because they could provide nonobvious insights in studying patterns of biological phenomena.

5.2. Benefits of collective categorization

The collective categorization method allows individuals to offer their own perspectives when describing a particular set of information. Section 4.1 showed the types of differences that could exist when people describe or group complex and ambiguous information such as biological mechanisms. The flexibility of collective categorization is particularly relevant for categorizing biological information.

Allowing this flexibility means that different people could describe the same information in multiple ways. In other words, semantic ambiguity can arise without explicit definitions of tags or categories, which poses challenges for computational reasoning.

However, this research has demonstrated the potential of collective categorization in discovering relationships between biological concepts. Since semantics involves the study of relations between concepts, our collective categorization method could help reveal the semantics of biological information. Halpin et al. (2007) and Robu et al. (2009) support that collaborative tagging could be analyzed to derive the meanings of the tags used.

The semantics identified from collective categorization could help discover relations that exist in biology. In addition, the results of collective categorization from multiple people could be compared with the modeling frameworks specified by researchers, to evaluate the consistency of frameworks.

Cheong et al. (2011) showed that a similar set of semantic relations as those reported in this paper could be used to identify biologically meaningful keywords (See Section 2.2.1.). Therefore, the semantics that emerge from collective categorization could also assist in the information-extraction process, giving clues to what kind of information should be captured from biological text.

5.3. Application of computational approach in collective categorization

To discover more relationships between biological concepts, more examples of biological transformation and participant categorization results could be incorporated. This can be resource-intensive because the data acquisition and categorization/tagging tasks are currently performed manually.

Natural-language processing techniques could automate or semi-automate both identification and categorization of examples. Descriptions of biological transformation may include verbs that are causally related, e.g., one verb describing the enabling mechanism and the other verb describing the desired action achieved. In such a case, the automatic extraction tool developed by Cheong and Shu (2012) could be used to identify candidate biological information.

For categorization, automatic tagging could be performed based on the keywords that appear in descriptions of the examples. Multiple tagging results could be populated by focusing on different part-of-speeches of keywords, i.e., verbs, nouns, adjectives, and mimic the process of multiple people tagging the examples with different perspectives.

The main benefit of a computational approach is that a large number of examples can be analyzed, allowing for the identification of more patterns of biological phenomena.

6. SUMMARY AND CONCLUSION

This work confirmed our suspicions that many challenges exist in collecting and categorizing biological information for the purpose of biomimetic design. The participants of this study used varying approaches to categorize/tag examples of biological transformations, resulting in different groups and attributes describing the same sets of examples.

A collective categorization method was developed to quantify the relatedness between groups that participants created. This method demonstrated the potential to identify some meaningful semantic relations between biological concepts, i.e., hierarchical, synonymous, and causal.

Aggregating multiple sets of free-form categorization and tagging could reveal semantics that can contribute towards developing formal representations or learning unique patterns in biological phenomena.

ACKNOWLEDGMENTS

REFERENCES

Altshuller G, 1984, Creativity as an Exact Science, Gordon & Breach, New York, NY.

Baclawski K, Niu T, 2006, Ontologies for Bioinformatics, MIT Press, Cambridge MA.

Camburn BA, Guillemette J, Crawford RH, Wood KL, Jensen DJ, Wood JJ, 2010, When to transform? Development of indicators for design context evaluation, Proc. ASME IDETC/CIE, Montreal, Canada, Aug. 15-18, DETC2010-28951.

Cheong H, Chiu I, Shu LH, 2010, Extraction and transfer of biological analogies for creative concept generation. Proc. ASME IDETC/CIE, Montreal, Canada, Aug. 15-18, DETC2010-86680.

Cheong H, Chiu I, Shu LH, Stone R, McAdams D, 2011, Biologically meaningful keywords for functional terms of the functional basis, Journal of Mechanical Design 133:021007.

Cheong H, Shu LH, 2012, Automatic extraction of causally related functions from natural-language text for biomimetic design, Proc. ASME IDETC/CIE, Chicago, IL, Aug. 12-15, DETC2012-70732.

Chiu I, Shu LH, 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.

Chakrabarti A, Sarkar P, Leelavathamma B, Nataraju B, 2005, A functional representation for aiding biomimetic and artificial inspiration of new ideas, Artificial Intelligence for Engineering Design Analysis and Manufacturing, 19(2):113-132.

Goel A, 1997, Design, analogy, and creativity, IEEE Expert, 123:62-70.

Goel A, Rugaber S, Vattam S, 2009, Structure, behavior, and function of complex systems: The structure, behavior, and function modeling language, Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 23(1):23-35.

Halpin H, Robu V, Shepherd H, 2007, The complex dynamics of collaborative tagging, *WWW 2007,* May 8–12, 2007, Banff, Alberta, Canada.

Han L, Chan G, 2007, HFCT: A hybrid fuzzy clustering method for collaborative tagging, International Conference on Convergence Information Technology.

Helms M, Vattam S, Goel A, 2009, Biologically inspired design: Process and products, Design Studies, 30(5):606-622.

Hirnschall H, 1975, Eyes on the Wilderness, Hancock House Pub Ltd, Saanichton, BC.

Ke J, Chiu I, Wallace JS, Shu LH, 2010, Supporting biomimetic design by embedding metadata in natural-language corpora, Proc. ASME IDETC/CIE, Montreal, Canada, Aug. 15-18, DETC2010-29057.

Kuhr R, Wood K, Jensen D, Crawford R, 2010, Concept opportunity diagrams, a visual modeling method to find multifunctional design concepts, Proc. ASME IDETC/CIE, Montreal, Canada, Aug. 15-18, DETC2010-29068.

Mak TW, Shu LH, 2004, Abstraction of biological analogies for design, CIRP Annals 53(1):117-120.

Mak TW, Shu LH, 2008, Using descriptions of biological phenomena for idea generation, Research in Engineering Design 19(1):21-28.

Nagel J, Nagel R, Stone R, McAdams D, 2010, Function-based, biologically inspired concept generation, Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 24(4):521-535.

Purves WK, Sadava D, Orians GH, Heller HC, 2011, Life, The Science of Biology 6/e. Sinauer Associates, Sunderland MA.

Robu V, Halpin H, Shepherd H, 2009, Emergence of consensus and shared vocabularies in collaborative tagging systems, ACM Transactions on the Web, 3(4):No.14.

Sadava D, Heller HC, Orians GH, Purves WK, Hillis DM, 2008, Life: The Science of Biology, 8th edition, Sinauer Associates Inc., Sunderland, MA.

Shu LH, 2010, A natural-language approach to biomimetic design, Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 24(4):483-505.

Singh V, Skiles SM, Krager JE, Wood KL, Jensen D, Sierakowski R, 2009, Innovations in design through transformation: A fundamental study of transformation principles, Journal of Mechanical Design 131:081010-1.

Umeda Y, Tomiyama T, 1997, Functional reasoning in design, IEEE Expert, 123:42-48.

Vandervenne D, Verhaegen P-A, Dewulf S, Duflou JR, 2011, A scalable approach for the integration of large knowledge repositories in the biologically-inspired design process, International Conference on Engineering Design, Technical University of Denmark, August 15-18.

Vincent M, 2002, Systematic technology transfer from biology to engineering, Phil. Trans. R. Soc. Lond. 360(1791):159-173.

Vincent J, Bogatyreva O, Bogatyrev N, Bowyer A, Pahl A, 2006, Biomimetics: Its practice and theory, Journal of the Royal Society, 3(9):471-48.