Knowledge Extraction and Annotation for Cross-Domain Textual Case-Based Reasoning in Biologically Inspired Design

Spencer Rugaber, Shruti Bhati[(✉)], Vedanuj Goswami, Evangelia Spiliopoulou,
Sasha Azad, Sridevi Koushik, Rishikesh Kulkarni, Mithun Kumble,
Sriya Sarathy, and Ashok Goel

School of Interactive Computing,
College of Computing Georgia Institute of Technology,
Atlanta, GA 30308, USA
Shruti.bhati@gatech.edu

Abstract. Biologically inspired design (BID) is a methodology for designing technological systems by analogy to designs of biological systems. Given that knowledge of many biological systems is available mostly in the form of textual documents, the question becomes how can we extract design knowledge about biological systems from textual documents for potential use in designing engineering systems? In earlier work, we described how annotating biology articles with partial Structure-Behavior-Function models helps users access documents relevant to a given design problem and understand the biological systems for potential transfer of their causal mechanisms to engineering problems. In this paper, we present an automated technique instantiated in the IBID system for extracting partial SBF models of biological systems from their natural language documents for potential use in biologically inspired design.

1 Background, Motivations and Goals

Biologically inspired design is a well-known design paradigm that uses nature as a source of practical, efficient and sustainable solutions to stimulate design of technological systems (Benyus 1997; Vincent and Mann 2002). Recently biologically inspired design has grown into a movement with an increasing number of engineering and system designers looking towards nature as a source of ideas (Lepora et al. 2013). Biologically inspired design entails cross-domain analogies: It views nature as a library of design cases and biologically inspired design as a process of abstracting, transferring and adapting designs of biological systems into designs of technological systems.

Biologically inspired design is also related to Textual Case Based Reasoning (TCBR) because knowledge of many biological systems is available mainly in the form of textual documents. Textual information typically is hard to process by

computers due to its relatively unstructured format and the numerous possible variations in its interpretation. Hence, it is often necessary to introduce additional processing to extract structured knowledge from textual documents. As a result, TCBR is commonly combined with techniques from information retrieval, natural language processing, text mining, and knowledge discovery (Weber et al. 2005). The question then becomes how can we extract design knowledge about biological systems from textual documents for potential use in designing engineering systems?

In earlier work, we observed that when engaged in biologically inspired design, design teams typically searched the Web for biology articles describing systems that might inspire solutions to their problems (Vattam and Goel 2013a). We also found that the design teams typically struggled to locate biology articles relevant to their problems because search engines are not designed specifically for cross-domain retrieval, and, in particular, keywords that describe biology articles do not capture the design semantics of the biological systems described in the articles. Thus, in earlier work (Vattam and Goel 2013b), we presented an interactive system called Biologue that annotated biology articles with partial Structure-Behavior-Function (SBF) models (Goel et al. 2009) of the biological systems described in the articles. We found that the SBF annotations on the biology articles enhanced the precision and relevance of retrieved articles.

However, the semantic annotations in Biologue were handcrafted, which raised the issues of scalability and repeatability. If we are to make the interactive retrieval not only relevant and precise but also scalable and repeatable, then we must develop computational techniques for automatically extracting the partial SBF models of biological systems from their natural language descriptions. The objective of this paper is to describe a preliminary, high-level computational process for extracting structures, behaviors and functions of biological systems from textual documents. This process is embodied in the Intelligent Biologically Inspired Design (IBID) system presently under development.

2   The Problem: An Illustrative Example

Consider an engineer interested in improving water harvesting for a village in an arid region. Suppose that the engineer seeks inspiration from nature. Darkling beetles that live in the Namib Desert, one of the hottest places on Earth, survive by using their shells to draw water from periodic fog-laden winds. Thus, two beetle species from the genus *Onymacris* have been observed to fog-bask on the ridges of sand dunes (Norgaard and Danke 2010). How might our hypothetical engineer find biology articles describing the fog-harvesting processes of the beetles? How might the engineer confirm that the retrieved descriptions are relevant to her design problem? How might the engineer build a deep enough understanding of the beetles' fog-harvesting mechanisms to support application to her problem?

The engineer might conduct a literature survey using a web search engine. However, the current search technology for conducting this kind of literature

survey is plagued by several problems (Vattam and Goel 2013a). First, biologically inspired design by definition entails cross-domain knowledge transfer from biology systems to technological designs. However, current search engines are not designed to support cross-domain search. Second, engineers and biologists speak different languages, and most engineers are novices in biology. This makes it difficult for engineers to interpret a biology article or even to form an effective query. Third, current search engines use keywords to filter their results. However, the keywords do not capture a deep understanding of the user's query of the design problem. Thus, the keyword-based search typically results in imprecise results, including voluminous hits on unrelated documents.

Even when a search engine notes an appropriate article, at best it highlights contents words that match keywords in the query. The engineer must still expend effort to understand the article sufficiently to determine relevance. Hence, there are two opportunities to improve engineer productivity: increase the precision of the set of retrieved articles and facilitate relevance checking by improved annotation.

3   Our Approach to Developing a Solution

IBID uses a representation for complex systems called Structure-Behavior-Function models (Goel et al. 2009). SBF models consist of three main parts. The Function submodel of a system is an abstraction over the system's actions on its external environment. The Structure sub-model expresses its physical components and the connections among them. The Behavior sub-model describes the causal mechanisms that arise from the interactions among the structural components and that accomplish the system's functions. We have previously used SBF models extensively in building theories of analogical design. In particular, Goel and Bhatta (2004) showed that domain-specific SBF models can be abstracted into Behavior-Function design patterns for cross-domain analogical transfer. We have also developed several tools to support biologically inspired design including DANE, a library of SBF models of biological systems (Goel et al. 2011) and Biologue (Vattam and Goel 2013b) mentioned above.

In a preprocessing phase in IBID, SBF models of biological systems are extracted from articles describing them, and the articles are annotated by the extracted models. While the functions in the extracted SBF models are expressed in a domain-independent controlled vocabulary, the structural components are expressed in a biology-specific vocabulary. Thus, in the current version of IBID, design queries are made by specifying the desired functions (in the domain-independent controlled vocabulary of functions), possibly augmented with a specification of biology-specific structural components. Biology articles are retrieved based on the match with the functional and structural annotations on the articles, thereby increasing precision. The retrieved articles are annotated with SBF model elements, thereby making it easier to evaluate the relevance of articles.

The extracted SBF model of the biological case contains pointers back into the document from which it was extracted for each of its model elements. When

the document is retrieved, the pointers can be used to annotate how a segment of text contributes to the model. For example, a biological process described in an article can have a function, such as *transport*, highlighted in yellow, and the object of transport, *water*, highlighted in green. Thus the engineer reading the article can more quickly determine the relevance of the article.

3.1   Example Continued

In order to elaborate on the analysis and extraction of the aforementioned SBF models, we consider the article (Norgaard and Danke 2010). Following is a text snippet from the article:

> "The mechanism by which fog water forms into large droplets on a beaded surface has been described from the study of the elytra of beetles from the genus *Stenocara*. The structures behind this process are believed to be hydrophilic peaks surrounded by hydrophobic areas; water carried by the fog settles on the hydrophilic peaks of the smooth bumps on the elytra of the beetle and form fast-growing droplets that - once large enough to move against the wind - roll down towards the head."

IBID identifies the following structural component from the snippet:

Structure:

– Name: elytra
– Properties: hydrophobic, hydrophilic, smooth
– Parts: grooves

Here is an example of one behavior extracted from this snippet.

Behavior:

– Predicate - move
– Cause - "water carried by the fog settles on the hydrophilic peaks of the smooth bumps on the elytra of the beetle and form fast-growing droplets that - once large enough"
– Effect - "roll down towards the head"

4   IBID

IBID is a web application that retrieves and annotates biology articles in support of BID. IBID uses SBF models and controlled vocabularies to facilitate its retrieval and annotation. Several other aspects of IBID are worth noting.

– **Natural language processing (NLP):** When it analyzes an article, IBID makes use of common NLP technology including parsing, part-of-speech tagging, and word-sense-disambiguation, to detect salient sections of the document. Technical vocabulary is detected by use of one or more domain-dependent taxonomies. For example, to analyze a document about the water-harvesting behavior of beetles, biological taxonomies from the fields of entomology and morphology might be used.

- **Taxonomies of Structures, Behaviors and Functions:** For the IBID system, we developed a domain-independent taxonomy of functions (Spiliopoulou et al. 2015). Our function ontology combines elements of the function ontology in the Functional Basis (Hirtz et al. 2002) with elements in AskNature's function ontology for biomimicry (Deldin et al. 2002). This is important because the domain-independence of the function taxonomy enables a cross-domain matching between functions delivered by biological systems and the functions desired in engineering problems. IBID also uses Vincent's (Vincent 2014) biology-specific taxonomy of structural components and connections. Finally, IBID uses a subset of Khoo et al.'s taxonomy of behavioral patterns (Khoo et al. 1998, 2000). These taxonomies of structures, behaviors and functions play a role in IBID similar to that of Schank's (1972) conceptual dependency in semantic processing: They help generate top-down expectations for completing the schemas corresponding to the elements in the taxonomies.
- **Semantic annotation:** Using the vocabulary and relations present in the SBF taxonomy, the textual content of a document can be semantically annotated. *Semantic annotation* is a technique that helps to add semantics to unstructured documents (Davies et al. 2006). In particular, IBID makes use of VerbNet (Kipper et al. 2008), a knowledge base of common verbs and their expected role-fillers. VerbNet further improves IBID's word-sense disambiguation, and its frames serve as the first level of IBID's semantic processing. For example, VerbNet was used to determine the roles (Names, Properties, and Parts) used in the Structure frame for elytra presented in the last section.
- **Faceted search:** When engineers query IBID's repository of biology articles, they do so using faceted search (Prieto-Diaz 1991). A *faceted search* interface provides an orthogonal set of controlled vocabularies, one for each dimension of the search space. These include the expected *title*, *author*, and *publication date* dimensions. More important, however, for achieving precision, is its use of dimensions for *structure*, *behavior* and *function*. For example, by selecting *water* as a structural element, the engineer can focus her search on specific kinds of biological processes.

The following subsections describe IBID's system architecture, computational process, data model, use cases and current status. The section concludes by relating the IBID approach to CBR.

4.1   IBID System Architecture

IBID uses a classical client–server architecture. The web client uses dynamic HTML, CSS and Javascript to support user query construction and perusal of results. The server is written in PHP, with analysis performed by a Java servlet. Extracted knowledge is stored in a MySQL database. In addition to the search-and-perusal scenario, IBID also supports two other uses cases: file upload and analysis and taxonomy management.

In Fig. 1, the user interacts with the interface of the tool via a web browser. The PHP server acts as an intermediary between the client side and the Java
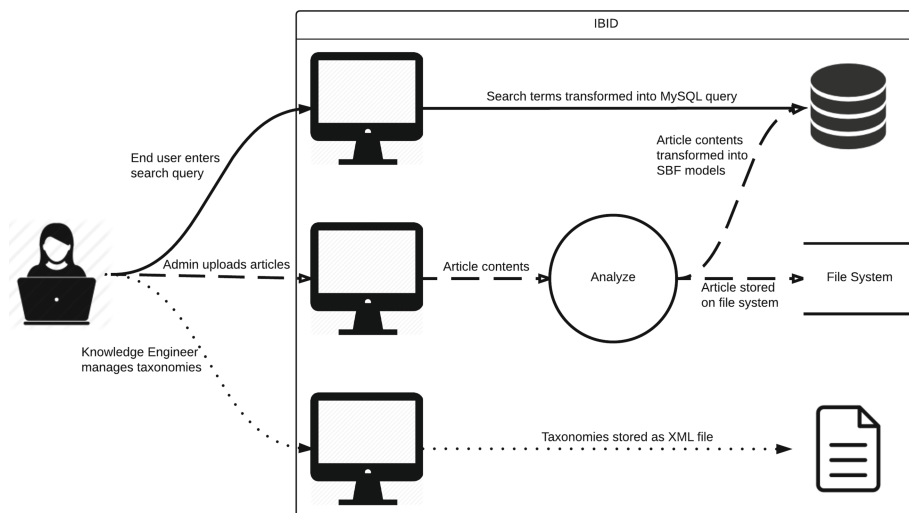
Fig. 1. IBID's conceptual architecture

semantic analyzer. The bold line in the figure depicts the user searching for models. The dashed line represents a scenario where an administrator can upload and analyze articles. The dotted line depicts a knowledge engineer managing taxonomies.

Both text and PDF files can be uploaded. Each file uploaded is sent to the Java analyzer. The file is parsed, and semantically processed to produce Structure-Behavior-Function models that are then stored into the MySQL database.

The core Java analyzer uses a number of NLP techniques to extract an SBF model of each biological process described in the article being analyzed. First, the article is broken down into sentences, each of which is parsed. The parse graph thus obtained is further processed by different modules to extract the Structure, Behavior and Function sub-models.

4.2   IBID's Computational Process for Extracting SBF Models

Figure 2 illustrates IBID's computational processes for extracting partial SBF models of biological systems from natural language documents. Initially, IBID breaks down the input text file into individual sentences and uses the Stanford parser (De Marneffe et al. 2011) to generate a parse tree for each.

Function Extraction: Function extraction focuses on the predicates present in each sentence. Using VerbNet's application programming interface (API), one or more frames are constructed for each sentence. The most relevant frame is selected and used to populate the SBF Function sub-model. The predicates
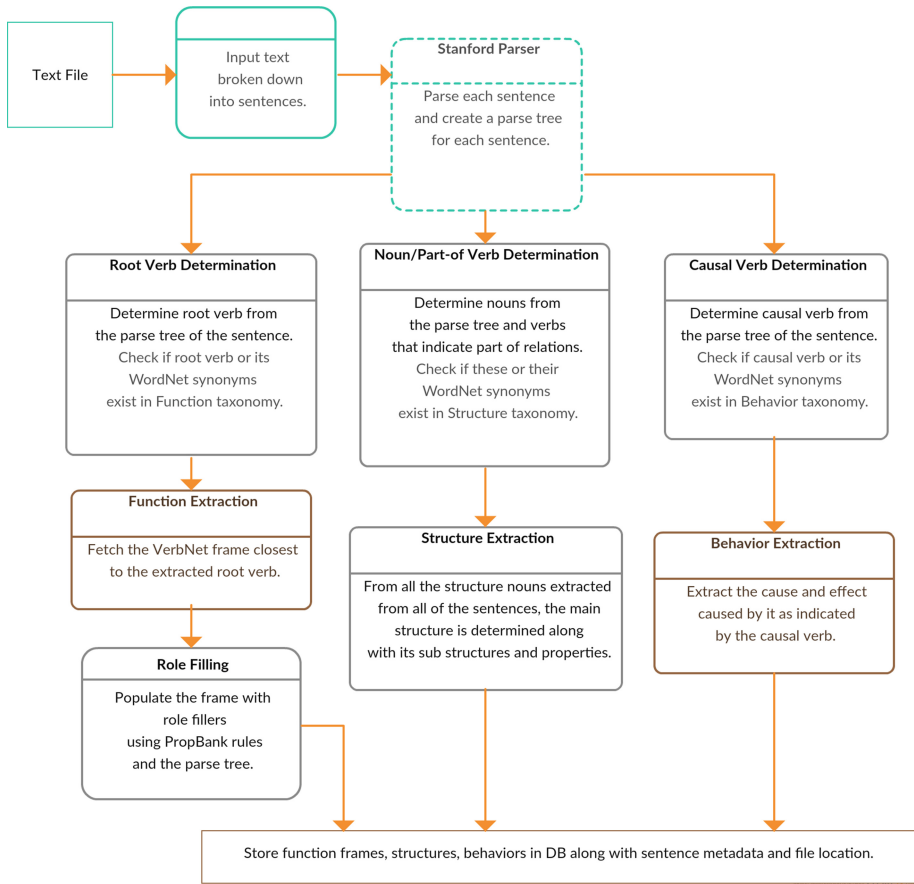
Fig. 2. IBID's computational processes for extracting SBF models of biological systems

selected have to be part of IBID's Function taxonomy. If this is not the case, a bespoke algorithm is applied to find the nearest match.

Behavior Extraction: The behavior of the biological system is captured in the form of causal chains: actions, effects and their causes. The action is the predicate in the sentence. The sentence is then matched to compiled patterns for causal chains to determine whether or not one is present.

Structure Extraction: Using the functional root verb in each sentence, the related subject and object are determined. Using WordNet (Miller 1995; Fellbaum 1998), synonyms and hypernyms are mapped into an ontology of biological components and connections due to Vincent (2014); only matches above a preset but tunable threshold are considered for further processing. The nouns

thus found are designated as structural components of the SBF Structure sub-model.

While the extracted structures, behaviors and functions are composed from multiple sentences in the input text, the extracted SBF models are at least partially domain specific. In particular, while the extracted functions are expressed in the domain-independent control vocabulary of functions, the extracted structures are expressed in a biology-specific vocabulary of structural components and connections. Thus, the user must do additional processing to extract domain-independent behavior-function patterns (Goel and Bhatta 2004) for transfer to engineering design problems.

4.3 IBID's Data Model

Figure 3 is a detailed enhanced entity–relationship diagram of IBID's data model. There are three groups into which all the tables have been arranged: articles, taxonomies and models. The group on the right corresponds to the tables related to articles. While the actual document contents are stored in the file system, IBID retains key information in its database (metadata and unique IDs) that are used during retrieval. The group at the bottom consists of three tables that store the taxonomies for the SBF sub-models. The group of tables to the left contains the stored models. As part of document analysis, IBID extracts SBF function, structure and behavior sub-models. The function information is stored in a format similar to VerbNet's frames. Structure is stored in a custom format in the structure entity table, and the causal behavior information is stored in the causality table.
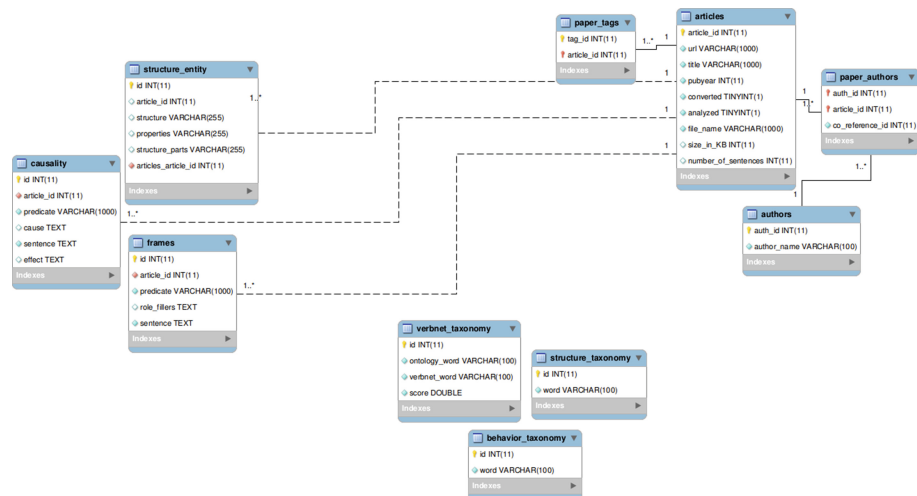
Fig. 3. Enhanced Entity-Relationship diagram for IBID

4.4   IBID Use Cases

The IBID tool has three main use cases: faceted search performed for an end-user engineer, document upload and analysis performed by a system administrator, and taxonomy management performed by a knowledge engineer.

Fig. 4 shows the first use case, the retrieval of several documents in which a biological process is accomplished via movement.
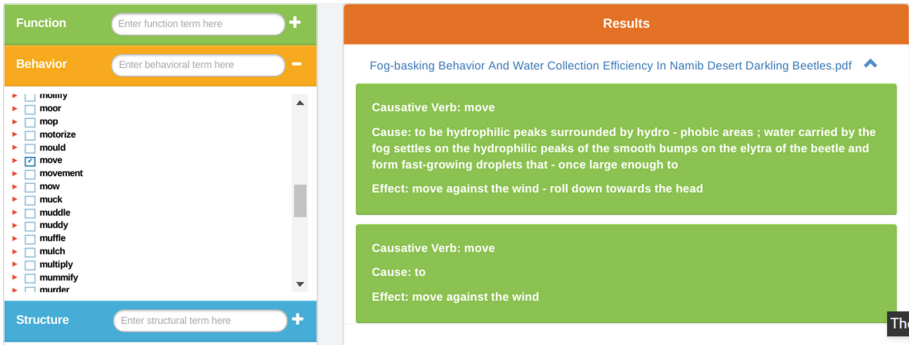

Fig. 4. Screenshot showing the Behavior cause and effects related to *move*.

When a user clicks on any term in the menu on the left side, the system expands its search and adds more synonyms and hypernyms for the term. All of these terms are then searched for in the database, and links to the documents that contain a related model are returned. The user can click on one of the links to peruse the document. The system highlights the relevant portion of the document to make it easier for the user to understand the results.

4.5   Current Status

IBID is a working prototype with all the above mentioned features implemented. Knowledge engineers can configure domain-specific vocabularies. System administrators can upload documents and analyze them. The analyzed articles are then tagged with SBF model elements. Once these models are stored, an engineer can search for and retrieve matching documents.

4.6   IBID and CBR

IBID serves as case based system in two ways. First, during the analysis phase, IBID extracts SBF models from documents. It uses the models as indices into its repository. It also searches for similar SBF models already in the repository. If similar cases exist, IBID stores the biological processes using the same SBF index.

Second, IBID acts as a CBR system during user search. By extracting SBF models for various biological processes, IBID indexes these processes. Using these SBF model indexes, IBID treats the various biological processes as cases. The cases stored by IBID can be retrieved using the index by searching either for Structures, Behaviors or Functions. The search results show the biological process that consist of the searched structure/behavior/function. This is the case-retrieval phase. From there, based on the biological processes retrieved, the user can adapt the process to her engineering problem taking advantage of her increased understanding of how the biological process works.

5   Validation

Our strategy for validating IBID has three main parts: (i) reliance on past work, (ii) execution of IBID on a large corpus of biology articles and inspection of results, and (iii) comparison with human performance.

5.1   Strategy 1: Reliance on Past Work

IBID assumes the following:

– Biologically Inspired Design is an effective design technique (Vincent and Mann 2002)
– SBF is robust in representing mechanisms in engineering and biology
– SBF models can improve search effectiveness for use in Biologically Inspired Design

5.2   Strategy 2: Execution of IBID on a Corpus of Articles

IBID's taxonomy of functions contain 8 functions at the top level of the hierarchy, with about 50 functions in all, with more than 45,000 hypernyms/synonyms. Its taxonomy of structural components and connections contains more than 200 elements. Thus, IBID is not a small system.

The IBID corpus of biology articles contains 255 journal papers and is a superset of Biolgue's corpus. We were able to successfully execute IBID on all biology articles in its corpus. Manual inspection of the SBF models extracted by IBID indicates that the models are incomplete but not incorrect. In particular, one way in which the SBF models are incomplete is that at present IBID does not fully relate the extracted structures, behaviors and functions with one another. For example, some of the structural elements it extracts do not appear to play any role in the accomplishment of the system functions. On the other hand, IBID presently does not always extract all the behaviors it should. Thus, IBID provides an automated computational technique for abstracting only partial SBF models of biological systems described in textual documents, storing the partial SBF models as biological design cases, and indexing the case by both their functions and structural elements in support of interactive biologically inspired design.

5.3   Strategy 3: Comparison with Human Performance

Our third strategy for ongoing evaluation of IBID focuses on comparing the quality of the SBF models extracted by IBID with those extracted by human experts. Ideally, the model extracted by IBID should be equivalent to the SBF model extracted by a human expert (a criteria that only a few practical CBR systems meet for tasks as complex as automated construction of a case library). Thus, we measured the Cohen's Kappa (Cohen 1960; Landis and Koch 1977) coefficient pairwise for models produced by a group of human evaluators. Preliminary Kappa results indicated an agreement of 0.559. A closer inspection uncovered three explanations:

First, there is no unique "best" SBF model for a complex biological system. There are always differences among SBF models generated by human experts as well, even when the human evaluators identify the same mechanisms. Thus, it is not easy to define and measure the degree of similarity between two SBF models, given that the same mechanism may be described in a different way in the models.

Second, for the above reason the Kappa coefficient is not the best measure for measuring the quality of SBF models extracted by IBID. In order to resolve these problems, we now use the Weighted Kappa Coefficient (Cohen 1968) that weighs each part of the model according to the reviewers' agreement on that part. Thus, when a word is described as a function by all the reviewers it is weighted more than when only half of the reviewers agree. Those weights are used later in order to calculate the similarity between IBID's extracted model and the models that humans' extracted.

Finally, as mentioned above, IBID extracts only partial SBF models: It does not presently fully integrate the structures, the behaviors and functions into a complete SBF model. We expect that the quality of the extracted SBF models to improve once IBID starts exploiting the constraints that full integration of SBF models will impose on decisions about individual structures, behaviors and functions.

6   Related Work

The IBID project relates to efforts in case-based reasoning, natural language processing, biologically inspired design, and computational creativity. In research on textual case-based reasoning, Weber et al. (2001) propose a knowledge management framework for acquiring cases from human experts as well as natural language documents. Bruninghaus and Ashley (1998, 2006) describe a technique for predicting the outcome of a legal case given a brief textual summary of the case facts. Schumacher et al. (2012) present a technique for extracting procedural knowledge from natural language documents available on the web. Sizov et al. in (2014, 2015) describe a technique for extracting causal relational graphs from natural language documents. Our work is related to the above research. The behaviors in SBF models can be viewed as graphs representing causal processes;

it differs from earlier work in that (i) IBID extracts multiple kinds of knowledge (structures, behaviors, functions) and (ii) it extracts SBF models for cross-domain analogical transfer from biological systems to engineering design.

In research on natural language processing, Berant et al. (2014) developed a system that answers multiple choice questions based on natural language paragraphs describing biological processes. Although their representation of causal processes is similar to that of IBID, their system uses manually preprocessed questions and answers. In research on biologically inspired design, Cheong and Shu (2012) have used natural language processing techniques to extract and categorize causally related biological functions. Finally, in research on computational creativity, Jursic et al. (2012) describe a process to identify and explore terms that relate different domains. In IBID, the taxonomy of functions provides the cross-domain words that lead to knowledge transfer from biology to engineering.

7   Conclusion

IBID is an interactive system for finding and semantically annotating biology articles relevant to a design problem. In a preprocessing phase, IBID extracts partial SBF models of biological systems from biology articles and uses the SBF models as annotations on the biology articles. Then, when the user specifies particular design functions of interest, IBID retrieves both the matching SBF models and the relevant biology articles. Thus, the ontology of functions acts as a cross-domain bridge between biological systems and engineering problems.

Work on IBID faces several types of challenges including disambiguating different senses of a word describing a function, a behavior or a structure; distinguishing between biological systems and other processes described in an article; improving the quality of extracted SBF models to match that of manually extracted models; and using the SBF models for supporting case-based reasoning in biologically inspired design. As mentioned earlier, IBID at present extracts only partial SBF models in that it does not fully integrate the structures, the behaviors and functions into a complete SBF model. The quality of the extracted SBF models should improve when IBID starts exploiting the constraints that full integration of SBF models will impose on decisions about individual structures, behaviors and functions.

References

Benyus, J.: Innovation Inspired by Nature. William Morrow, New York (1997)
Vincent, J.F., Mann, D.L.: Systematic technology transfer from biology to engineering. Philos. Trans. Roy. Soc. Lond. A Math. Phys. Eng. Sci. **360**(1791), 159–173 (2002)

Lepora, N.F., Verschure, P., Prescott, T.J.: The state of the art in biomimetics. Bioinspir. Biomim. **8**(1), 013001 (2013)

Weber, R.O., Ashley, K., Bruninghaus, S.: Textual case-based reasoning. Knowl. Eng. Rev. **20**(03), 255–260 (2005)

Vattam, S.S., Goel, A.K.: An information foraging model of interactive analogical retrieval. In: Proceedings of 35th Annual Meeting of the Cognitive Science Society (2013a)

Vattam, S.S., Goel, A.K.: Biological solutions for engineering problems: a study in cross-domain textual case-based reasoning. In: Proceedings of 21st International Conference on Case-Based Reasoning, pp. 343–357 (2013b)

Goel, A.K., Rugaber, S., Vattam, S.: Structure, behavior, and function of complex systems: the structure, behavior, and function modeling language. Artif. Intell. Eng. Des. Anal. Manufact. **23**(01), 23–35 (2009)

Norgaard, T., Dacke, M.: Fog-basking behaviour and water collection efficiency in Namib Desert Darkling beetles. Front. Zool. **7**(1), 1 (2010)

Goel, A., Bhatta, S.: Use of design patterns in analogy-based design. Adv. Eng. Inf. **18**(2), 85–94 (2004)

Goel, A.K., Vattam, S., Wiltgen, B., Helms, M.: Cognitive, collaborative, conceptual and creative - four characteristics of the next generation of knowledge-based cad systems: a study in biologically inspired design. Comput. Aided Des. **44**(10), 879–900 (2011)

Spiliopoulou, E., Rugaber, S., Goel, A., Chen, L., Wiltgen, B., Jagannathan, A.K.: Intelligent search for biologically inspired design. In: Proceedings of the 20th International Conference on Intelligent User Interfaces Companion, pp. 77–80. ACM, March 2015

Hirtz, J., Stone, R., McAdams, D., Szykman, S., Wood, K.: A functional basis for engineering design: reconciling and evolving previous efforts. Res. Eng. Des. **13**(2), 65–82 (2002)

Deldin, J.M., Schuknecht, M.: The AskNature database: enabling solutions in biomimetic design. In: Goel, A., McAdams, D., Stone, R. (eds.) Biologically Inspired Design, pp. 17–27. Springer, London (2014)

Vincent, J.F.: An ontology of biomimetics. In: Goel, A., McAdams, D., Stone, R. (eds.) Biologically Inspired Design: Computational Methods and Tools, pp. 269–285. Springer, London (2014)

Khoo, C., Kornfit, J., Oddy, R.N., Myaeng, S.H.: Automatic extraction of cause-effect information from newspaper text without knowledge-based inferencing. Lit. Linguist. Comput. **13**, 177–186 (1998)

Khoo, C.S.G., Chan, S., Niu, Y.: Extracting causal knowledge from a medical database using graphical patterns. In: Proceedings of 38th Annual Meeting of the Association for Computational Linguistics (ACL2000), pp. 336–343 (2000)

Schank, R.: Conceptual dependence: a theory of natural language understanding. Cogn. Psychol. **3**(4), 532–631 (1972)

Davies, J., Studer, R., Warren, P.: Semantic Web Technologies: Trends and Research in Ontology-Based Systems. Wiley, New York (2006)

Kipper, K., Korhonen, A., Ryant, N., Palmer, M.: A large-scale classification of English verbs. Lang. Resour. Eval. **42**(1), 21–40 (2008)

Prieto-Diaz, R.: Implementing faceted classification for software reuse. Commun. ACM **34**(5), 88–97 (1991)

De Marneffe, M.C., MacCartney, B., Manning, C.: Generating typed dependency parses from phrase structure parses. In: Proceedings of European Language Resources and Evaluation Conference (LREC 2006) (2006)

Miller, G.A.: WordNet: a lexical database for English. Commun. ACM **38**(11), 39–41 (1995)

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

Cohen, J.: A coefficient for agreement of nominal scales. Educ. Psychol. Meas. **20**(1), 37–46 (1960)

Landis, J., Koch, G.: The measurement of observer agreement for categorical data. Biometrics **33**(1), 159–174 (1977)

Cohen, J.: Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. Psychol. Bull. **70**(4), 213 (1968)

Weber, R., Aha, D.W., Sandhu, N., Munoz-Avila, H.: A textual case-based reasoning framework for knowledge management applications. In: Proceedings of the 9th German Workshop on Case-Based Reasoning. Shaker Verlag, pp. 244–253 (2001)

Bruninghaus, S., Ashley, K.: Developing mapping and evaluation techniques for textual case-based reasoning. In: Workshop Notes of the AAAI 1998 Workshop on Textual CBR (1998)

Bruninghaus, S., Ashley, K.D.: Progress in textual case-based reasoning: predicting the outcome of legal cases from text. In: Proceedings of the National Conference on Artificial Intelligence (Vol. 21, No. 2, p. 1577). Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press, 1999, July 2006

Schumacher, P., Minor, M., Walter, K.: Extraction of procedural knowledge from the web: a comparison of two workflow extraction approaches. In: Proceedings of the 21st International Conference on World Wide Web, pp. 739–747 (2012)

Sizov, G., Ozturk, P., Styrac, J.: Acquisition and reuse of reasoning knowledge from textual cases for automated analysis. In: Lamontagne, L., Plaza, E. (eds.) ICCBR 2014. LNCS, vol. 8765, pp. 465–479. Springer, Heidelberg (2014)

Sizov, G., Öztürk, P., Aamodt, A.: Evidence-driven retrieval in textual CBR: bridging the gap between retrieval and reuse. In: Hüllermeier, E., Minor, M. (eds.) ICCBR 2015. LNCS (LNAI), vol. 9343, pp. 351–365. Springer, Heidelberg (2015). doi:10. 1007/978-3-319-24586-7_24

Berant, J., Srikumar, V., Chen, P.C., Vander Linden, A., Harding, B., Huang, B., Manning, C.D.: Modeling biological processes for reading comprehension. In: EMNLP, October 2014

Cheong, H., Shu, L.H.: Automatic extraction of causally related functions from natural-language text for biomimetic design. In: ASME 2012 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, pp. 373–382. American Society of Mechanical Engineers, August 2012

Jursic, M., Cestnik, B., Urbancic, T., Lavrac, N.: Cross-domain literature mining: Finding bridging concepts with CrossBee. In: Proceedings of the 3rd International Conference on Computational Creativity, pp. 33–40, May 2012