Towards a Virtual Librarian for Biologically Inspired Design – Knowledge-Based Methods for Document Understanding

Ruth Petit-Bois[1], Jeffrey Jacob[2], Spencer Rugaber[3], Ashok Goel[4]
Design & Intelligence Lab, School of Interactive Computing, Georgia Institute of Technology[1,2,4]
School of Computer Science, Georgia Institute of Technology[3]
petitbois@gatech.edu[1], jeffrey.jacob@gatech.edu[2], rugaber@cc.gatech.edu[3], goel@cc.gatech.edu[4]

Abstract

IBID is a virtual librarian that processes biology articles and builds semantic annotations based on the contents of an article. It then assists human designers by locating and presenting biology articles related to a design query. IBID's use of ontologies allows for knowledge extraction and assists users with the identification of key information in an article and comparison of the contents of two articles. In this paper, we describe how the addition of an environment ontology enhances IBID's capability to understand the habitats of various organisms. In a pilot study, we evaluated IBID's performance against human subjects who read the same passage and highlighted phrases pertaining to locations and habitats. The preliminary results indicate that the ability to add ontologies to IBID allows it to extract meaning from new documents.

1 Introduction

Scientific documents are information-rich and are more common and more available than ever before. However, with this proliferation comes the challenge of tracking and understanding scientific documents at scale. Traditionally, a scientist could work with a librarian to find the literature relevant to the problem of interest. Now, most scientific literature has moved online, real librarians are hard to find, and it is increasingly difficult, even for experts, to track, read and understand all the new scientific documents that are being generated on a given topic.

Understanding scientific documents is an involved process: there is a big difference between just reading text and actually understanding it. We view scientific document understanding as the ability to process information and then be able to draw useful inferences from it and not draw spurious inferences. This view supports higher level tasks like comparing the contents of two different documents and identifying similarities and differences between them.

We posit that AI can be a powerful ally in tracking and understanding scientific documents and that knowledge-based methods that use ontologies can augment the understanding capability of AI agents. This kind of AI agent can serve as a sort of virtual librarian for scientific literature. The IBID (Intelligent Biologically Inspired Design; Goel et al. 2020; Rugaber et al. 2016) interactive system is intended to be a virtual librarian for the domain of biologically inspired design in which designers of technological systems look to the natural world for ideas (Goel 2013a; Goel, McAdams & Stone 2014). In this paper, we describe how IBID's use of ontologies allows for knowledge extraction and can assist users with tasks like identifying key information in an article and comparing the contents of two different articles. In particular, we show how the addition of an environment ontology enhances IBID's capability to understand the locations and habitats of various organisms.

2 Related Research

Biologically inspired design, also known as biomimicry (Beynus 1997) and as biomimetics (Vincent & Mann 2002) is a paradigm for sustainable and environmentally friendly design. Consider, for example, the Namib Desert Beetle: The insect survives in the acrid desert by harvesting fog droplets that stick to its wings (Naidu and Hattingh 1988). If engineers could successfully and efficiently mimic this ability in technological systems at scale, it might be possible to solve many water crises that exist in the world (Chen & Zhang 2020).

However, there are several major hurdles in putting biologically inspired design paradigms into practice. From an information-processing perspective, one big hurdle is locating biological cases relevant to a design problem. Given a

---

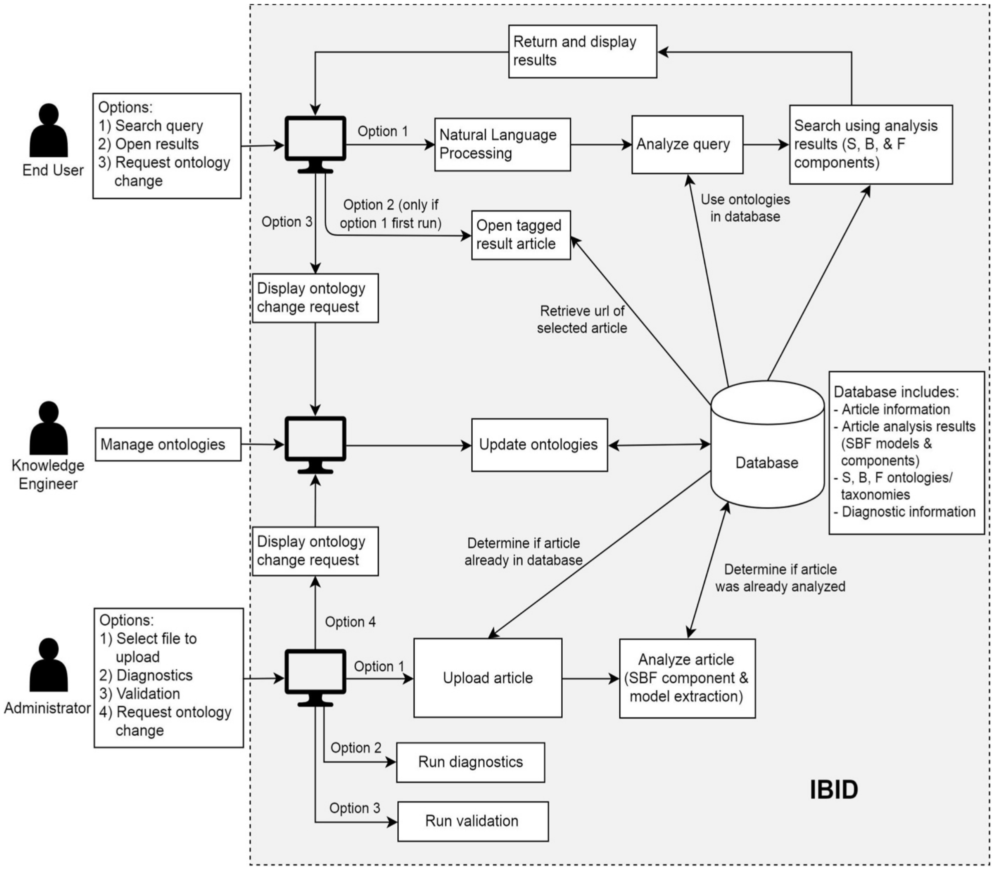


Figure 1. The Conceptual Architecture of IBID

problem, most designers search for articles describing relevant biological systems online. Observations of online information-seeking behavior of (student) designers indicate three problems (Vattam & Goel 2011, 2013): *Findability* – designers have difficulty finding biology articles relevant to a design problem; *Recognizability* – designers have difficulty recognizing that an article describes a biological system that is relevant to their problem; and *Understandability* – designers have difficulty understanding the biological system described in an article.

As a result, there have been several attempts in using natural language processing techniques to help designers locate biology articles relevant to their problem. Shu (2010) describes an early approach in engineering for using natural language processing for this task. Shu uses keywords for anchoring the natural language processing, but points out that the benefits of information extraction through natural language processing is not restricted to known patterns. Nagel, Stone & McAdams (2010) use an engineering to biology thesaurus that translates design queries in engineering to equivalent keywords in biology. Krupier et. al (2017) provide a more recent effort coming from biology. Their work is based on a domain-specific ontology of biological systems (Vincent 2014) and focuses on identifying inter-relations in biological systems.

Of course, the goal of using publicly available scientific literature to support human creativity extends far beyond the domain of biologically inspired design. In the context of computational creativity more generally, Abgaz et al. (2017) use natural language processing to find analogies between constructs in research papers on computer graphics, and Lavrac et al. (2019) describe text mining techniques for detecting bridging concepts between seemingly unrelated terms in different articles such as *migraine* and *magnesium*.

3 Intelligent Biologically Inspired Design

The goal of the IBID project is to address the above mentioned problems of findability, recognizability and understandability in the context of biologically inspired design. Figure 1 shows the full functionality of IBID for its three use cases: (1) End users such as engineers and designers looking for biology articles relevant to their design problems, (2) Knowledge engineers extending IBID's knowledge representation ontologies, and (3) System administrators adding to its repository of analyzed papers. Figure 1 also specifies the actions available to each user type;

the arrows in the figure indicate progression of steps and/or access to/from the database.

The core of IBID's approach to these problems is the use of the Structure-Behavior-Function (SBF) models of technological and natural systems (Goel 2013b; Goel, Rugaber & Vatttam 2009) that originate from Chandrasekaran's Functional Representation scheme (Chandrasekaran 1994; Chandrasekaran, Goel & Iwasaki 1993). By an ontology we mean the specification of concepts and their relationships to other concepts (Chandrasekaran, Josephson & Benjamins 1999; Guarino, Oberle & Staab 2009). The SBF model of a system, technological or natural, is based on an ontology composed of several subontologies:

- *Structure Ontology:* The components, elements, or substances in a biological process.
- *Behavior Ontology:* The causal mechanisms or the processes of a biological system.
- *Function Ontology:* The outcome, result or the purpose of a biological systems.
- *Ontology of Relationships:* Relationships between structure and behavior and between behavior and function.

In earlier work on the KA project in the 1990s (Goel et. al 1996), we showed how an AI agent could learn an SBF model of a new device (such as a shaving cream can) from its natural language description in *The Way Things Work* (Macaulay 1988) by adapting the SBF model of a similar device (such as a fire extinguisher) stored in the agent's memory. More recently, we have shown that manually annotating biology articles by SBF models enhance their findability and recognizability (Vattam & Goel 2011) and as well as their understandability (Helms, Vattam & Goel 2010). IBID seeks to automatically extract the SBF models of the biological systems described in the articles.

3.1 Structure-Behavior-Function Ontologies

In IBID, the SBF ontologies come from several sources:

- *Structure Ontology* is borrowed from Vincent's (2014) ontology of biological systems.
- *Behavior Ontology* builds on Khoo et al.'s (1998) patterns of cause and effect.
- *Function Ontology* was developed in our laboratory (Rugaber et al. 2016). Functional concepts are organized hierarchically: similar concepts are grouped together as families and more nuanced concepts are found deeper in the hierarchy.

The current version of IBID does not directly relate structure, behaviors and functions of a biological system into a complete SBF model.

These ontologies help IBID construct a partial SBF model of the biological system described in a biology article. Rugaber et. al. (2016) provide an example of how IBID processes the following passage from Norgaard and Dacke (2010):

> 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 processes the above paragraph and identifies the structure, behavior and function specified in it:

- *Structure:* IBID identifies the entity in question as *elytra*.
- *Behavior:* IBID identifies the cause as *droplets grow in size* and the effect as they *roll down towards the head.*
- *Function:* IBID identifies the result of the action as *move* the water droplets.

This list is only illustrative of IBID's capabilities, not comprehensive. IBID performs this kind of automatic extraction of structure, behavior and function for whole articles and annotates the articles with the extracted structure, behaviors and functions.

Given IBID's annotation of biology articles in a corpus with the structure, behaviors and functions of biological systems described in them, users can perform faceted search on the corpus (Prieto-Diaz 1991). Thus, a user may search for the function *move,* or the structural element *elytra,* or both. A user may also use IBID to perform a search using a design query expressed in plain English: given such a query, IBID extracts the structure, behaviors and functions of the desired technological system from the query and then matches the extractions with the SBF annotations on the articles in the corpus in a manner similar to the earlier KA project (Peterson, Mahesh & Goel 1994). This helps IBID address the problems of findability and recognizability we described earlier.

IBID also highlights the SBF annotations on a biology article. This helps IBID address the problem of understandability even for dense and long articles, such as the Norgaard and Dacke (2010) article quoted above. This can potentially help the user process biology articles more efficiently and easily, where the users may include not only biologists, but also engineers, designers, or even citizen scientists.

4 Adding an Environment Ontology to IBID

While the paragraph from the Norgaard and Dacke (2010) article briefly mentions the location of *elytra (elytra* of *beetles*), the above description of IBID has no way of identifying the location of the structural elements of a biological system. However, for many biological systems, the location, habitat, and, more generally, the environment of the system is very important. The external environment is also important for technological systems: the specification of many design problems includes a specification of the environment of the desired technological system (Helms & Goel 2014). Thus, there is a need to add an environment ontology so that IBID can identify the locations and habitats of biological systems.

Actually, the environment always was a part of SBF modeling (Goel 2013b). For example, Prabhakar & Goel (1998) analyzed the functioning of technological systems such as a room air conditioning system not only in terms of its structure, behaviors and functions, but also its external environment. The research question for the IBID project is whether we can add an environment ontology to the SBF ontology and if IBID can use the new ontology to identify the locations and habitats of biological systems just as it identifies their structures, behaviors and functions.

Instead of building a new environment ontology from scratch, we decided to explore already existing ontologies. After examining several candidates, we selected the Environment Ontology (ENVO) described by Buttigieg et al. (2013) in the *Journal of Biomedical Semantics*. This ontology is hosted on the OBO Foundry (Smith et. al. 2007) and is quite comprehensive. A big advantage of this ontology over many others is that it can be exported as a Web Ontology Language (OWL) file. OWL files written in the Semantic Web Language are "designed to represent rich and complex knowledge about things, groups of things, and relations between things." (McGuinness & Van Harmelen 2004; Web Ontology Language at www.w3.org/OWL/). Given that the OWL file containing ENVO was developed by highly skilled biologists, it eliminated the need for us to spend time creating the links between concepts manually. Not only are the links already made, but ENVO is made up of hundreds of nodes of concept names, definitions, parents, synonyms, notes, and other metadata that describe ecosystems, entire planets, and other astronomical bodies, and their parts (Buttigieg et. al. 2013). Integrating new knowledge into IBID is efficient and easy because of the use of OWL files and sourcing them from places like OBO Foundry helps IBID leverage the knowledge of domain experts.

To provide a simpler testing ground of adding an ontology into IBID and testing its effectiveness, we reduced ENVO to just contain extremely basic concepts relating to ecosystems and their key environmental concepts. Figure 2 illustrates a small excerpt from the stripped-down ENVO.

4.1 Generalizing the Approach

Three factors were especially important in adding the ENVO ontology to IBID:

1. The ability to have a standard format by which to import ontologies.
2. The ability to add information quickly without breaking previous implementations.

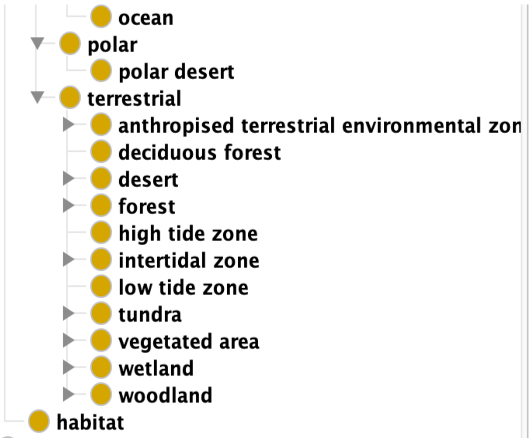


Figure 2. An excerpt of stripped-down ENVO.

3. The ability to use & export this information easily.

By using ENVO, it was clear that Objective 1 could be reached just by establishing that all future ontologies would use the OWL format. Not only can OWL files be imported, parsed, modified, and exported easily, there are many tools to help visualize and act on these OWL files such as Protégé (Musen 2015). Protégé became the software used to scale down ENVO, as well as rebuild the Structure, Behavior, and Function ontologies so they also conformed to the new OWL standard. Adding new concepts or modifying existing concepts was simple using the Protégé software, thereby addressing Objective 2.

With the new converted ontologies, the issue of how to store these ontologies in a relational database arose. To resolve this, we developed a script that would take in an OWL file and convert it into its relational database equivalent. By the end of the implementation, the structure, behavior, and function ontologies were updated and reimported into IBID's relational database using the new OWL format. The environment ontology was also imported into IBID allowing for articles to be analyzed to extract environment concepts. In addition to this, IBID now has a pipeline for integrating new ontologies that are in the OWL format in an easy manner. Given that all of the data was imported into a relational database, exporting this information from the database was simple, and even using Protégé to export the OWL files into other formats was simple, addressing Objective 3.

5 Experimentation

With the ability to import new knowledge executed, the next step was experimenting and evaluating how well IBID could leverage this knowledge. An experiment was conducted to test the effectiveness of the environment ontology with ten participants outside of the IBID project in the Design and Intelligence Lab. In conjunction with this experiment, a validation page was developed to test the functionality of the environment ontology. The use case of comparing scientific documents was also explored qualitatively.

5.1 Validation of Environment Ontology

IBID's validation took a passage of text and ran it through IBID's analysis pipeline and returned a list of results specific to the environment ontology. The experiment compared IBID's results with human subjects analyzing the same passages. The results of this experiment would reveal gaps in the environment ontology's functionality that could be used to make it more robust. The text for the experiments came from Szalay (2014), en.wikipedia.org/wiki/Elephant, and McTighe (2011).

In total, 10 human participants completed the experiment. Each participant read the same three passages on three different organisms. The instructions were to underline terms in the passages they considered to be related to the "environment" or the "habitat" in which organisms live. The organisms in question were the King Cobra, with a passage containing 4 sentences, the African Bush Elephant with a passage containing 14 sentences, and the Highland Streaked Tenrac with a passage containing 6 sentences.

Passage	**Avg. # of Terms by Humans**	**# of Terms by IBID**
King Cobra (4 sentences) – Passage 1	8	2
African Bush Elephant (14 sentences) – Passage 2	10	4
Highland Streaked Tenrac (6 sentences) – Passage 3	11	4

Table 1. Results from the first pass of the experiment

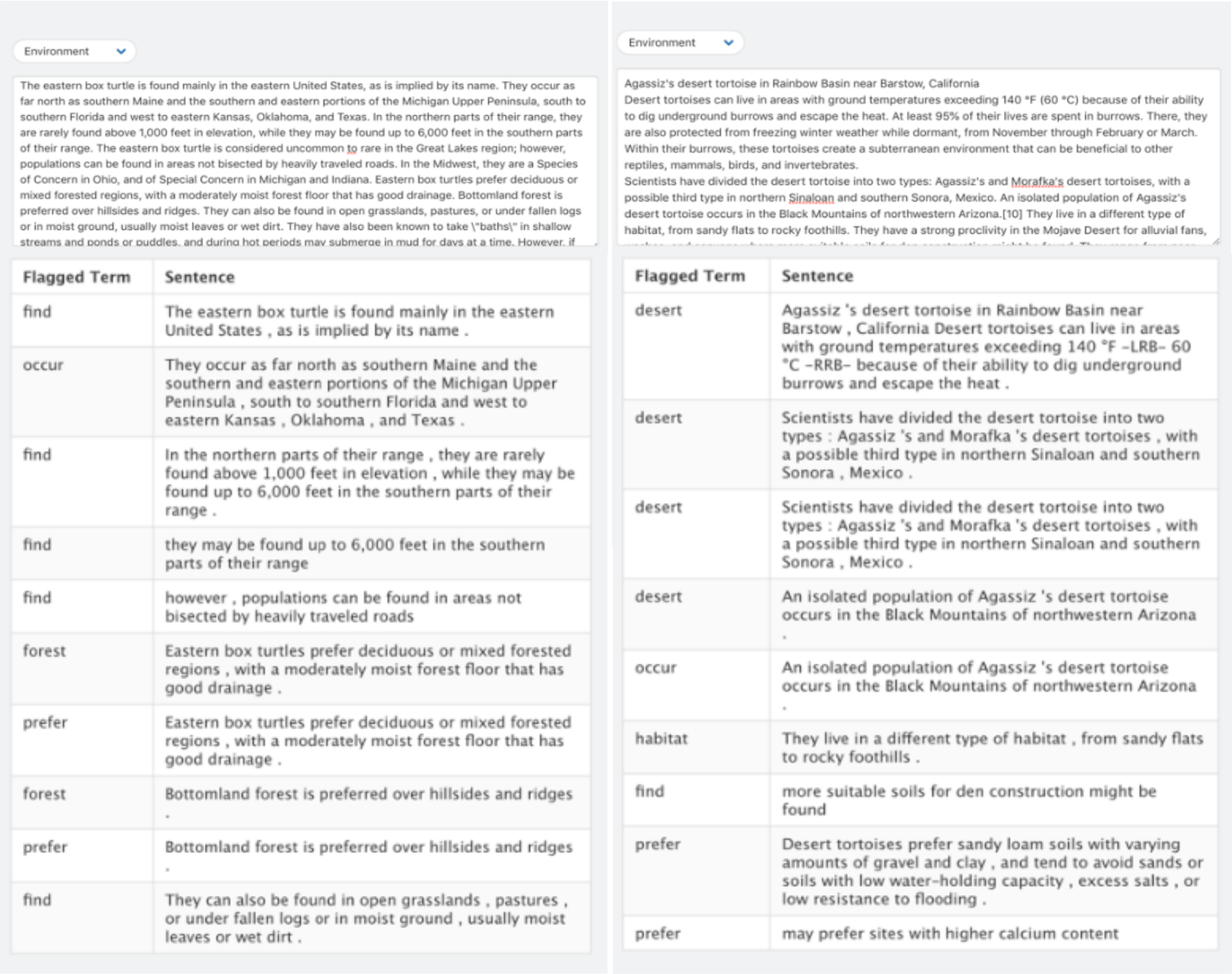
The highlighted phrases were pulled out exactly as they were marked by the participant. The assumption here was that there was a difference between a term having been highlighted in one straight stroke, and a term being highlighted with spaces in between. This meant that in this sentence from en.wikipedia.org/wiki/Elephant:

> The African bush elephant can be found in habitats as diverse as dry savannahs, deserts, marshes, and lake shores, and in elevations from sea level to mountain areas above the snow line.

There was a difference if a participant highlighted, "*dry savannahs, deserts, marshes, and lake shores*" in one go to count as one phrase, or they highlighted "*dry savannahs,*" then "*deserts,*" then "*marshes,*" then "*lake shores*" separately to count as 4 different phrases. Of the 4 sentences based on the King Cobra's habitat, the humans were on average able to locate ~8 different environment terms. Running the same passage in IBID led to it finding only 2. Of the 14 sentences based on the African Bush Elephant, the humans were able to on average find 10 different environment terms; IBID was able to identify 4. Finally, the passage on the Highland Streaked Tenrac had humans denoting around 11 environment terms while IBID was able to extract 4. The results are shown in Table 1.

Sentence (where >50% of Users Agree)	**Passage #**	**Phrase Selected** (where at least 5 people agreed on the concept/phrase)
They prefer streams in dense or open forest, bamboo thickets, adjacent agricultural areas and dense mangrove swamps.	Passage 1	streams in dense or open forest(x6)
		bamboo thickets (x8)
		adjacent agricultural areas (x5)
		dense mangrove swamps(x5)
The African bush elephant can be found in habitats as diverse as dry savannahs, deserts, marshes, and lake shores, and in elevations from sea level to mountain areas above the snow line.	Passage 2	dry savannahs (x7)
		Deserts (x8)
		Marshes (x8)
		lake shores (x8)
Forest elephants mainly live in equatorial forests but will enter gallery forests and ecotones between forests and savannahs.	Passage 2	equatorial forests (x7)
Asian elephants prefer areas with a mix of grasses, low woody plants, and trees, primarily inhabiting dry thorn-scrub forests in southern India and Sri Lanka and evergreen forests in Malaya.	Passage 2	dry thorn-scrub forests (x5)
		evergreen forests (x7)
Elephants tend to stay near water sources.	Passage 2	stay near water sources (x6)
Highland streaked tenrecs are found in schlerophyllous and montane forests and adjacent areas at elevations of 1550 to 1800 m.	Passage 3	Schlerophyllous (x5)
		montane forests (x5)
They occur both in primary rainforests and in introduced forests of eucalyptus and pine.	Passage 3	primary rainforests (x6)
		introduced forests of eucalyptus and pine (x7)

Table 2. The aggregated results for the three passages.



The eastern box turtle is found mainly in the eastern United States, as is implied by its name. They occur as far north as southern Maine and the southern and eastern portions of the Michigan Upper Peninsula, south to southern Florida and west to eastern Kansas, Oklahoma, and Texas. In the northern parts of their range, they are rarely found above 1,000 feet in elevation, while they may be found up to 6,000 feet in the southern parts of their range. The eastern box turtle is considered uncommon to rare in the Great Lakes region; however, populations can be found in areas not bisected by heavily traveled roads. In the Midwest, they are a Species of Concern in Ohio, and of Special Concern in Michigan and Indiana. Eastern box turtles prefer deciduous or mixed forested regions, with a moderately moist forest floor that has good drainage. Bottomland forest is preferred over hillsides and ridges. They can also be found in open grasslands, pastures, or under fallen logs or in moist ground, usually moist leaves or wet dirt. They have also been known to take \"baths\" in shallow streams and ponds or puddles, and during hot periods may submerge in mud for days at a time. However, if

Flagged Term	Sentence
find	The eastern box turtle is found mainly in the eastern United States , as is implied by its name .
occur	They occur as far north as southern Maine and the southern and eastern portions of the Michigan Upper Peninsula , south to southern Florida and west to eastern Kansas , Oklahoma , and Texas .
find	In the northern parts of their range , they are rarely found above 1,000 feet in elevation , while they may be found up to 6,000 feet in the southern parts of their range .
find	they may be found up to 6,000 feet in the southern parts of their range
find	however , populations can be found in areas not bisected by heavily traveled roads
forest	Eastern box turtles prefer deciduous or mixed forested regions , with a moderately moist forest floor that has good drainage .
prefer	Eastern box turtles prefer deciduous or mixed forested regions , with a moderately moist forest floor that has good drainage .
forest	Bottomland forest is preferred over hillsides and ridges .
prefer	Bottomland forest is preferred over hillsides and ridges .
find	They can also be found in open grasslands , pastures , or under fallen logs or in moist ground , usually moist leaves or wet dirt .

Agassiz's desert tortoise in Rainbow Basin near Barstow, California
Desert tortoises can live in areas with ground temperatures exceeding 140 °F (60 °C) because of their ability to dig underground burrows and escape the heat. At least 95% of their lives are spent in burrows. There, they are also protected from freezing winter weather while dormant, from November through February or March. Within their burrows, these tortoises create a subterranean environment that can be beneficial to other reptiles, mammals, birds, and invertebrates.
Scientists have divided the desert tortoise into two types: Agassiz's and Morafka's desert tortoises, with a possible third type in northern Sinaloan and southern Sonora, Mexico. An isolated population of Agassiz's desert tortoise occurs in the Black Mountains of northwestern Arizona.[10] They live in a different type of habitat, from sandy flats to rocky foothills. They have a strong proclivity in the Mojave Desert for alluvial fans,

Flagged Term	Sentence
desert	Agassiz 's desert tortoise in Rainbow Basin near Barstow , California Desert tortoises can live in areas with ground temperatures exceeding 140 °F -LRB- 60 °C -RRB- because of their ability to dig underground burrows and escape the heat .
desert	Scientists have divided the desert tortoise into two types : Agassiz 's and Morafka 's desert tortoises , with a possible third type in northern Sinaloan and southern Sonora , Mexico .
desert	Scientists have divided the desert tortoise into two types : Agassiz 's and Morafka 's desert tortoises , with a possible third type in northern Sinaloan and southern Sonora , Mexico .
desert	An isolated population of Agassiz 's desert tortoise occurs in the Black Mountains of northwestern Arizona .
occur	An isolated population of Agassiz 's desert tortoise occurs in the Black Mountains of northwestern Arizona .
habitat	They live in a different type of habitat , from sandy flats to rocky foothills .
find	more suitable soils for den construction might be found
prefer	Desert tortoises prefer sandy loam soils with varying amounts of gravel and clay , and tend to avoid sands or soils with low water-holding capacity , excess salts , or low resistance to flooding .
prefer	may prefer sites with higher calcium content

Figure 3. Example of Comparing Results from Two Documents about an Eastern Box Turtle and Desert Tortoise.

Table 2 contains the concepts that a majority of participants agreed on. The criteria for "agreeing" means that of the aggregated list of results, at least 50% of the participants agreed that the selected sentence was one that contained an environment concept and at least 5 participants also agreed on the concept that indicated it related to the environment.

5.2 Comparing Two Documents

As mentioned earlier, scientific document understanding allows an agent to perform higher level tasks and one such task that is paramount in any kind of research is the ability to quickly compare the key points of two different documents. IBID is able to take in two documents and run its analysis and display the results side-by-side. This process involves the same pipeline as discussed earlier and leverages the same knowledge base. We tested this process with several different excerpts taken from descriptions of the habitats of different species, an example of which is shown in Figure 3. It can be seen that IBID's ability to understand the key concepts in a document helps the researcher quickly compare two documents. If we have two documents about the same or similar species, IBID can help the researcher compare and contrast information and see where two different documents are in agreement and where they disagree. We believe that this can be a powerful tool and a major feature in the realm of scientific document understanding.

6 Discussion and Results

Based on the experiment above, we can see that the addition of a new ontology, in this case the environment ontology, improves IBID's understanding in this domain. IBID initially had no understanding ability when it came to habitats and locations, but the addition of this ontology led to increased understanding as seen in Table 1. However, we acknowledge that the number of participants in our experiment was small and IBID did not reach human level performance. We still feel that these preliminary results show that

IBID's ability to integrate new knowledge moves it towards becoming a true virtual librarian.

The experiment also showed some of the weaknesses IBID has. For example, there are many proper noun *location* words (country names, cardinal directions, etc.) that many participants deemed relevant to the *environment* of an animal. IBID's knowledge base is strictly that of *habitats* as described in ENVO. Take for instance the simple sentence from Passage 3 (McTighe 2011):

> They are most commonly found at forest fringes on the central plateau edge and near cultivated fields and rice paddies

The key term was "forest" and it was pulled out by IBID; the term "forest" maps to an environment concept in ENVO. In contrast to this, humans are able to look at a sentence saying, "southern Indian desert" and see that the whole phrase indicates location while IBID would only be able to recognize the term "desert".

Looking at the "Phrase Selected" column in Table 2, it is clear that there are many examples where humans agreed that adjectives describing habitats are just as important as the habitat itself. Descriptive words like "dense mangrove swamps" and "dry savannahs" might be difficult for IBID to parse because they are compound terms containing a descriptive word followed by a habitat word. This issue could be addressed by extending IBID's parser to include adjectives that might describe an environment term.

One thing IBID does really well is identify the verb predicates from a sentence. Verbs like "prefer", "occur", and "find", occur frequently with environment related phrases that were marked by the human participants. For example, in Passage 3 (McTighe 2011), IBID identifies the phrase,

> tenrecs are found in schlerophyllous and montane forests and adjacent areas at elevations of 1550 to 1800 m.

where the verb used to identify this phrase is "find". Although the specific environment terms don't map to concepts in the ontology, IBID was able to extract this information.

There are sentences where IBID identified information that was right, but the term used to do so was not. For example, in the sentence from en.wikipedia.org/wiki/Elephant,

> The African bush elephant can be found in habitats as diverse as dry savannahs, deserts, marshes, and lake shores, and in elevations from sea level to mountain areas above the snow line.

IBID pulled out the word "bush" instead of one of the environment terms, even though "bush" is just part of the species name. This means that in passages where the name of an animal is an environment concept, IBID may pull out a false positive.

Finally, there were cases where humans identified vague habitat phrases like "under a tree," or "near water sources" that IBID missed. For example, in the sentence from en.wikipedia.org/wiki/Elephant:

> Asian elephants prefer areas with a mix of grasses, low woody plants, and trees, primarily inhabiting dry thorn-scrub forests in southern India and Sri Lanka and evergreen forests in Malaya.

It makes sense that humans marked "mix of grasses" and "low woody plants, and trees." However, there aren't any real concepts in ENVO that are mapped to by these phrases. However, the verb "prefer" was identified by IBID and allowed the sentence to be extracted independent of the environment terms found by the participants.

These results show that IBID's knowledge-based methods show promise in efficiently extracting information from a scientific document and that the use of ontologies allows for it to quickly integrate and leverage new knowledge, without the need for extensive data collection or training. Another major benefit of IBID's approach is better explainability. It is easy to determine gaps in IBID's knowledge, like those identified in regard to proper nouns and cardinal directions. It is also easy to see which knowledge IBID used to extract information. The use of an ontology also allows IBID and its users to leverage the relationships that are found for downstream inferencing tasks. The use of the standard OWL file format also allows users to edit the knowledge using tools like Protégé.

Conclusions

IBID demonstrates the effectiveness of knowledge-based methods in augmenting scientific document understanding and moves us towards a true virtual librarian. IBID's use of standardized ontologies allows it to quickly gain a deeper understanding of a new domain, without the need to acquire lots of new data or to spend time learning a complex model. This ability also allows IBID to be extensible. The Environment Ontology was a working example, but the same process can be applied to new ontologies, thus growing IBID's understanding capability. These abilities allow IBID to facilitate higher level tasks like document comparison, which can help users of IBID compare and contrast different approaches to their engineering problem. We acknowledge that there is a need for augmenting the analysis and filling the gaps in IBID's knowledge, but the use of knowledge-based methods helps the user to efficiently identify these gaps and easily make modifications or add extra processing.

Acknowledgements

We are grateful to Julian Vincent for sharing his ontology of biological systems; IBID's structure ontology is a subset of his structure ontology. We thank Pablo Boserman and Daniel Dias for their work on making the IBID system functional, as well as members of the Georgia Tech Design & Intelligence Lab for assisting with the IBID experiment including the evaluation of interactive system.

References

Abgaz, Y., Chaudhry, E., O'Donoghue, D., et al. 2017. Characteristics of pro-c analogies and blends between research publications. In Proceedings of the 8th International Conference on Computational Creativity, 1–8.

Beynus, J. 1997. *Biomimicry: Innovation Inspired by Nature.* New York: Harper Perennial. doi.org/10.1002/inst.12116

Buttigieg, P., Morrison, N., Smith, B., et al. (2013) The environment ontology: contextualising biological and biomedical entities. *Journal of Biomedical Semantics*, 4(1):43. doi.org/10.1186/2041-1480-4-43

Chandrasekaran, B. 1994. Functional Representation: A Brief Historical Perspective. *Applied AI*, 8(2): 173-197. doi.org/10.1080/08839519408945438

Chandrasekaran, B., Goel, A., & Iwasaki, Y. 1993. Functional representation as design rationale. *IEEE Computer*, 48-56. doi.org/10.1109/2.179157

Chandrasekaran, B., Josephson, J., Benjamins, V. 1999. What are ontologies and why do we need them? Intelligent Systems and their Applications. *IEEE Intelligent Systems,* 14(1):20–26. doi.org/10.1109/5254.747902

Chen, Z., & Zhang, Z. 2020. Recent progress in beetle-inspired superhydrophilic-superhydrophobic micropatterned water-collection materials. *Water Science & Technology 82 (2): pp. 207–226.* doi.org/10.2166/wst.2020.238

Goel, A. 2013a. Biologically Inspired Design: A New Program for Computational Sustainability, IEEE Intelligent Systems, 28(3), 80-84.

Goel, A. 2013b. One Thirty Year Long Case Study; Fifteen Principles: Implications of an AI Methodology for Functional Modeling. *Artificial Intelligence for Engineering Design Analysis and Manufacturing, 27(3): 203- 215, 2013.* doi.org/10.1017/S0890060413000218

Goel, A., Hagopian, K., Zhang, S., & Rugaber, S. 2020. Towards a Virtual Librarian for Biologically Inspired Design. In Proceedings of the 9th International Conference on Design Computing and Cognition, 377-396.

Goel, A., Mahesh, K., Peterson, J., & Eiselt, K. 1996. Unification of Language Understanding, Device Comprehension and Knowledge Acquisition. In Proceedings of the 10th Knowledge Acquisition for Knowledge-Based Systems Workshop, Banff, Canada.

Goel, A., McAdams, D., & Stone, R. 2014. *Biologically Inspired Design: Computational Methods and Tools*. London: Springer-Verlag. doi.org/10.1007/978-1-4471-5248-4

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.* doi.org/10.1017/S0890060409000080

Guarino, N.; Oberle, D.; & Staab, S. 2009. What is an ontology? Handbook on Ontologies, Edited by R. Studer. 1-17. Berlin: Springer-Verlag Berlin Heidelberg. doi.org/10.1007/978-3-540-92673-3_0

Helms, M., & Goel, A. 2014. The Four-Box Method: Problem Formulation and Analogy Evaluation in Biologically Inspired Design. *Journal of Mechanical Design, 136(11):111106.* doi.org/10.1115/1.4028172

Helms, M., Vattam, S., & Goel, A. 2010. The effects of functional modeling on understanding complex biological systems. In Proceedings of the ASME 2010 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Montreal: American Society of Mechanical Engineers, 107-115. doi.org/10.1115/DETC2010-28939

Khoo, C., Kornfilt, J., Oddy, R., & Myaeng, S-H. 1998. Automatic Extraction of Cause-Effect Information from Newspaper Text Without Knowledge-based Inferencing. Literary and Linguistic Computing, 13(4):177–186. doi.org/ 10.1093/llc/13.4.177

Kruiper, R.; Vincent, J.; Chen-Burger, J. & Desmulliez, M. 2017. Towards identifying biological research articles in computer-aided biomimetics. In Proceedings of the Conference on Biomimetic and Biohybrid Systems, 242–254. Springer. doi.org/10.1007/978-3-319-63537-8_21

Lavrac, N., Jursic, M., Sluban, et al. 2019. Bisociative Knowledge Discovery for Cross-domain Literature Mining. In *Computational Creativity*, edited by T. Veale & A. Cardoso, 121-139. Springer. doi.org/10.1007/978-3-319-43610-4_6

Macaulay, D. 1988. *The Way Things Work.* Boston, MA: Houghton Mifflin Company.

McGuinness, D., & Van Harmelen, F. 2004. OWL web ontology language overview. *W3C recommendation*, 10(10).

McTighe, L. 2011. Hemicentetes nigriceps. Available at animaldiversity.org/accounts/Hemicentetes_nigriceps/

Musen, M.A. 2015. The Protégé project: A look back and a look forward. *AI Matters.* Association of Computing Machinery Specific Interest Group in Artificial Intelligence, 1(4): 4-12. doi.org/10.1145/2757001.2757003

Nagel, J., Stone, R., & McAdams, D. 2010. An engineering-to-biology thesaurus for engineering design. In Proceedings of the ASME 2010 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Montreal: American Society of Mechanical Engineers. doi.org/10.1115/DETC2010-28233

Naidu, S., & Hattingh, J. 1988. Water Balance and Osmoregulation in Physadesmia Globosa, a Diurnal Tenebrionid Beetle from the Namib Desert. *Journal of Insect Physiology* 34(10): 911-917. doi.org/10.1016/0022-1910(88)90126-6

Norgaard, T., & Dacke, M. 2010. Fog-basking behavior and water collection efficiency in Namib Desert Darkling beetles. *Frontiers of Zoology* 7(1), 1. doi.org/10.1186/1742-9994-7-23

Peterson, J., Mahesh, K., Goel, A. 1994. Situating natural language understanding within experience-based design. *International Journal of Human-Computer Studies* 41(6): 881-913. doi.org/10.1006/ijhc.1994.1085

Prabhakar, S., & Goel, A. 1998. Functional modeling for enabling adaptive design of devices for new environments. *AI in Engineering 12(4): 417–444.* doi.org/10.1016/S0954-1810(98)00003-X

Prieto-Diaz, R. 1991. Implementing faceted classification for software reuse. *Communications of the ACM, 34(5): 88-97.* doi.org/10.1145/103167.103176

Rugaber, S., Bhati, S., Goswami, et al. 2016. Knowledge Extraction and Annotation for Cross-Domain Textual Case-Based Reasoning in Biologically Inspired Design. In Proceedings of the 24th International Conference in Case- Based Reasoning, 342-355. doi.org/10.1007/978-3-319-47096-2_23

Shu, L. 2010. A Natural-language approach to biomimetic design. *Artificial Intelligence for Engineering Design Analysis and Manufacturing,* 24(4):507–519. doi.org/10.1017/S0890060410000363

Smith, B.; Ashburner, M.; Rosse, C. et al. 2007. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, 25(11):1251–1255. doi.org/10.1038/nbt1346

Szalay, J. 2014. Facts about Cobras.. Live Science. 19 December, 2014. Available at https://www.livescience.com/43520-cobra-facts.html.

Vattam, S., & Goel, A. 2011. Foraging for inspiration: Understanding and supporting the information seeking practices of biologically inspired designers. In Proceedings of the ASME 2011 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. doi.org/ 10.1115/DETC2011-48238

Vattam, S., & Goel, A. 2013. Seeking Bioinspiration Outline: A Descriptive Account. In Proceedings of the 19th International Conference on Engineering Design, 517-526.

Vincent, J. 2014. An ontology of biomimetics. In *Biologically Inspired Design: Computational Methods and Tools*, edited by A. Goel, D. McAdams & R. Stone, 269-285, London: Springer. doi.org/10.1007/978-1-4471-5248-4_11

Vincent, J., & Mann, D. 2002. Systematic technology transfer from biology to engineering. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences,* 360(1791): 159-173. doi.org/10.1098/rsta.2001.0923

Web Ontology Language. 2011. By the OWL Working Group. Available from www.w3.org/OWL/. Published 11 December 2011.