# BIOINSPIRED MATERIALS DESIGN: A TEXT MINING APPROACH TO DETERMINING DESIGN PRINCIPLES OF BIOLOGICAL MATERIALS

A Dissertation

by

JOANNA NIFON TSENN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

| | |
|---|---|
| Chair of Committee, | Daniel McAdams |
| Committee Members, | Julie Linsey |
| | Richard Malak |
| | Douglas Allaire |
| | Sarbajit Banerjee |
| Head of Department, | Andreas Polycarpou |

May 2016

Major Subject: Mechanical Engineering

## ABSTRACT

Biological materials are often more efficient and tend to have a wider range and combination of properties than present-day engineered materials. Despite the limited set of components, biological materials are able to achieve great diversity in their material properties by the arrangements of the material components, which form unique structures. The structure-property relationships are known as structural design principles. With the utilization of these design principles, materials designers can develop bioinspired engineered materials with similarly improved effectiveness. While considerable research has been conducted on biological materials, identifying beneficial structural design principles can be time-intensive. To aid materials designers, the research in this dissertation focuses on the development of a text mining algorithm that can quickly identify potential structural design principles of biological materials with respect to a chosen material property or combination of properties.

The development of the text mining tool involves four separate stages. The first stage centers on the creation of a basic information retrieval algorithm to extract passages describing property-specific structural design principles from a corpus of materials journal articles. Although the Stage 1 tool identifies over 90% of the principles (recall), only 32% of the returned passages are relevant (precision). The second stage investigates text classification techniques to refine the program in order to improve precision. The classic techniques of machine learning classifiers, statistical features, and part-of-speech analyses, are evaluated for effectiveness in sorting passages into relevant and irrelevant classes. In the third stage, manual identification of patterns in the returned passages is employed to create a rule-based method. The resulting Stage 3 algorithm's precision values increase to 45%. In the final stage of algorithm development, the manual rule-based classification method is revisited to identify stricter rules to further emphasize precision. The Stage 4 algorithm successfully improves overall precision to 65% and reduces the number of returned passages by 74%, which allows a materials designer to more quickly identify useful principles. Finally, the research concludes with a validation that the text mining tool effectively identifies structural design principles and that the principles can be used in the development of bioinspired materials.

## ACKNOWLEDGEMENTS

I would first like to thank my co-advisors, Dr. McAdams and Dr. Linsey, for their guidance and support through the entire course of this research. I would also like to thank my committee members, Dr. Malak, Dr. Allaire, and Dr. Banerjee, for their time and efforts. I want to extend my gratitude to the National Science Foundation (NSF EEC 1025155/1338383), which provided partial financial support for this work.

I am grateful to my lab mates and the friends that I have made at Texas A&M University for making my time here a great experience. I appreciate their advice and reassurance regarding my research, but am most thankful for the relief they provided from my research studies. A special thanks to Madison, Nicole, Elissa, Soo, Edgar, and Sean for going above and beyond.

Thanks also go to my parents and sister for their continual support and love.

Finally, a special thanks needs to be extended to my husband, Joseph, for his unwavering love, encouragement, and patience throughout the entire process.

# TABLE OF CONTENTS

## LIST OF FIGURES

Page

## LIST OF TABLES

# CHAPTER I

# INTRODUCTION

Bioinspired design uses nature as a source of inspiration for creating solutions to engineering design problems. Through billions of years of development, nature has evolved time-tested, efficient designs that may offer innovative solutions. For over 3,000 years, designers have looked to nature for inspiration, although only more recently has nature been considered in the design of materials.

Biological materials are efficiently designed to accomplish specific functions. Man-made ("engineered") materials have many fundamental differences from biological materials [1]. Materials engineers and designers can use inspiration from the component structuring of natural materials in order to create bioinspired engineered materials with improved, or even new, material properties. For example, the properties of strength and toughness are usually mutually exclusive within engineered materials, but bioinspired materials have been developed that can achieve that very property combination [2]. While there are a variety of bioinspired materials that have been developed, there is not a well-known method or tool to support a materials designer seeking bioinspiration.

The research presented in this dissertation describes the development of a descriptive text mining algorithm that can aid materials designers in quickly identifying biological materials structural design principles relevant to their design needs. Data mining is the process of extracting valuable information and discovering knowledge from large recorded data sets using a computer-based methodology, and text mining derives the knowledge from textual data [3].

The rest of the introduction chapter lays the foundation of this dissertation by further clarifying the notion of biological and bioinspired materials, and by presenting the motivation for the research.

## Introduction to Biological Materials

Biological materials are composites based on ductile organic components and brittle inorganic components. Often, biological materials contain a complex hierarchical structure of both organic and mineral components [4]. One major difference between biological materials and engineered materials is that biological materials have limited component options. The principal organic components consist of collagen, keratin, chitin, elastin, cellulose, and protein, while mineral components include hydroxyapatite, calcium carbonate, and silica [5, 6]. While the properties of the components themselves are typically poor, the overall biological materials' properties are amplified beyond the simple rule of mixtures formulation [7]. The large range of properties that biological materials encompass are due to the arrangement of the components into material structures [8, 9]. For example, tendon, ligament, skin, and cartilage are all primarily composed of elastin and collagen, but each of these materials have very different properties and functions. Tendon needs to transfer loads with minimal energy loss and stretching when dealing with small applied strains. Since tendon needs to be stiff, it has a greater amount of collagen, and the collagen fibers form ordered rope-like structures.

Ligament, on the other hand, needs to cope with large strains in order to stabilize bones, so it is composed of more elastin and has a less regular arrangement of the collagen fibers. Skin can achieve deformation in all directions without elongating individual fibers due to its wavy fishnet-like structure, while cartilage accomplishes its duty of shape recovery and flexibility with multiple layers laid at different orientations [9].

Another notable characteristic of biological materials is that many of them have a hierarchical structure that spans a variety of length scales. Hierarchical structures can be found in a variety of materials from glass sponges to trees to spider silk [8, 10-12]. This differs from most engineering materials which undergo processing to impose a new structure, leading to a material that is essentially homogeneous at all length scales [13]. In biological materials, small building blocks are used to structure components for the next hierarchical level, which can then be used as building blocks for the next level, and so on. This allows for the optimization of the material at different length scales and the creation of multifunctional materials [12, 13]. The smallest level is often at the nanometer scale, while the largest scale may be at the macroscale [14]. For example, the glass sponge Euplectella has at least seven levels of hierarchy, as seen in Figure 1 [10]. At the lowest level, it consists of hydrated silica nanospheres (Figure 1a), which are arranged in layers and separated by organic interlayers (Figure 1b), all of which are layered concentrically to form a lamellar spicule structure (Figure 1c). These silica spicules are part of a layered ceramic-fiber composite (Figure 1d) that is bundled to form a strut (Figure 1e). The struts form a rectangular lattice, which is where the structure reaches the macroscale (Figure 1f). Finally, the lattice is used to form the overall cylindrical structure (Figure 1g). There are a variety of structures from which inspiration can be drawn within a single biological material.

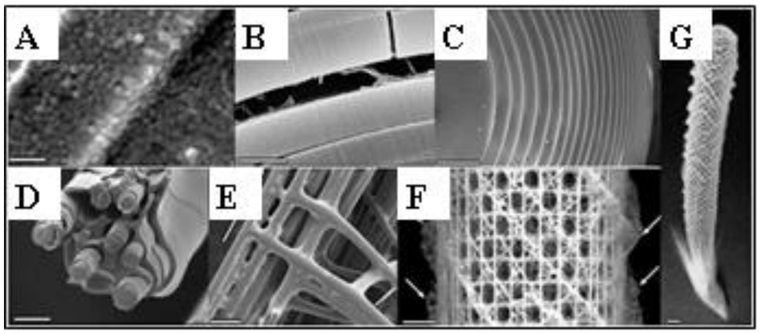



Figure 1. The hierarchical levels of Euplectella [10]: (a) Silica nanospheres, scale bar 500 nm. (b) Silica layers and organic interlayer, scale bar 1 μm. (c) Spicule consisting of concentric layers, scale bar 5 μm. (d) Beam formed from spicules in a silica matrix, scale bar 20 μm. (e) Struts consisting of bundled spicules, scale bar 100 μm. (f) Lattice of struts, scale bar 5 mm. (g) Entire skeleton, scale bar 1 cm.

**Introduction to Bioinspired Materials**

Bioinspired design uses nature as a source of inspiration to solve engineering design problems. Thus, when developing a bioinspired material, one looks to biological materials for reference. Natural materials are of interest because they are generally efficient in achieving their purposes, and their material properties can exceed those of the components by orders of magnitude [2]. However, it is difficult to copy a biological material exactly because the material grows and adapts in nature, a process which laboratory and manufacturing environments are mostly unable to replicate. Instead, the general strategies exhibited by these biological materials should be applied to bioinspired materials. Wegst and Ashby found that a bioinspired material is more successful if it draws inspiration from the biological material's structuring of the components rather than attempting to replicate the full hierarchical structure of the material or reproducing the chemistry and physiological functions [9]. For example, Bonderer et al. developed a bioinspired material based on nacre's brick-and-mortar structure. Nacre, also known as mother-of-pearl (Figure 2a), contains inorganic nanoplatelets and an organic matrix which form the brick-and-mortar structure, as seen in Figure 2b [15]. Rather than using the actual material components of nacre, the geometry of the platelets and mechanical properties of the components were mimicked in a layered hybrid film in order to develop a bioinspired material (Figure 2c) with a combination of high tensile strength and ductility, resulting in a full order of magnitude increase in toughness over nacre [16]. Similarly, materials design research has been conducted to improve other material properties and behavior by mimicking biological materials' underlying structural concepts for inspiration. This includes the development of an adhesive system mimicking gecko setae [17], a tough composite inspired by wood's arrangement of helically wound fibers to increase toughness [18], and a network polymer that mimics the sacrificial domains found in biopolymers in order to imitate the biopolymers' exceptional combination of strength and elasticity [19]. While bioinspired materials research is a relatively newer field, research in the area is rapidly growing, as displayed in Figure 3.

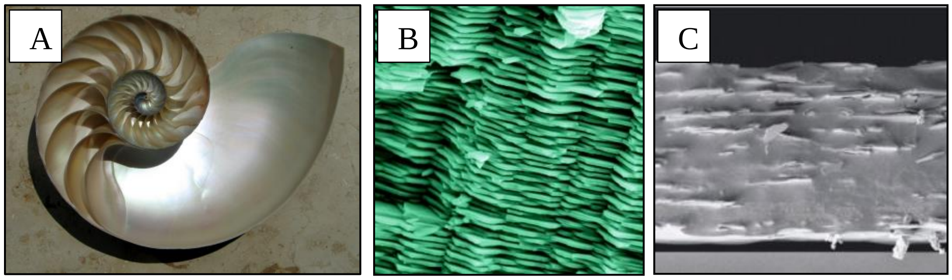



Figure 2. (a) Nacre can be found as an iridescent layer in shells [20]. (b) Nacre's brick-and-mortar structure [15]. (c) The bioinspired layered hybrid film [16].

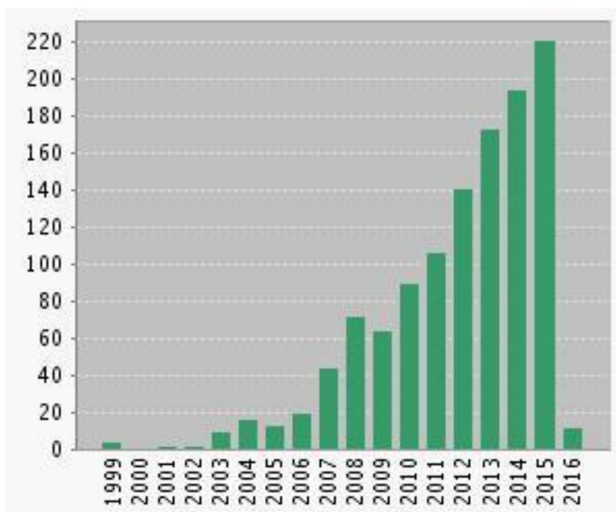



Figure 3. The number of published "bioinspired materials" items within the Web of Science Core Collection by year, as of February 6, 2016 [21].

## Motivation and Research Objective

Biological materials have structural patterns that commonly appear throughout the natural world. One such structure is the use of nanoplatelets embedded in a soft organic matrix, as found in enamel, dentin, vertebral bone, and nacre [14]. Another example is the utilization of cellular structures to balance strength, stiffness, and weight. These can be found at a variety of length scales, including the cell walls in parenchyma, bamboo stalks, wood, glass sponges, and bird beaks [22-25]. Squid beaks, teeth, and tendon-bone interfaces all use a gradient structure to alleviate modulus mismatch [26]. There are many other common structural design features found in biological materials beyond these few examples. These features are referred to as underlying "structural design principles". Because these archetypal structures appear in a variety of materials that evolved independently, it is hypothesized that they represent highly efficient solutions. The structural design principles can be used to design bioinspired materials by providing inspiration for material component structuring, rather than developing an entirely new material with different components [8, 27]. The different arrangements and structures of the material components affect the properties of the material, and these structure-property relationships can be applied at different length scales and for different functions.

Materials design is often conducted on a trial-and-error basis and requires designer expertise to predict the effects of component combinations and processing on the properties [28]. By approaching materials design from an engineering design standpoint, a more methodical design approach can be taken. Biological materials research is an ongoing field of investigation, which is useful because there is a plethora of inspirational sources for engineered materials. However, the amount of literature and knowledge available regarding biological material structures is now too much for a materials designer to read through to identify all of the potential structural design principles. A second challenge to manually identifying design principles is the interdisciplinary nature of the bioinspired materials field, which encompasses areas such as biology, materials science, chemical engineering, mechanical engineering, chemistry, genetics, veterinary medicine, and physics. Another issue is that a disconnect exists between biological materials research and materials designers. There are only a handful of survey papers that describe biological material structure-property relationships for bioinspired materials design [2, 29, 30]. There currently is not a well-known method that can support a materials designer seeking bioinspiration.

The goal of this research is to create a descriptive text mining algorithm that can automatically identify structural design principles with respect to a selected material property or combination of properties. The development of an algorithm helps to bridge the knowledge gap between biological materials research and materials designers. Rather than requiring a materials designer to read through full articles to extract useful information, the tool will allow the designer to sift through short passages. By reducing the amount of time required to identify structural design principles, a designer may be more inclined to look to natural solutions for inspiration. The overall goal is to develop an algorithm that is not just identifying a specific material system, but rather strategies for materials design. The identified structural design principles can then be used in the creation of novel bioinspired materials with properties tailored to their specific applications.

An example of a passage describing a structural design principle that may inspire a materials designer is:

> *"…AFM-based single molecule pulling experiments on the muscle protein titin revealed that unraveling of loosely packed modules led to increased molecular contour length and an increase in the energy absorbed prior to catastrophic failure. Similar sacrificial connective proteins have been proposed as toughening agents in the macroscopic mechanical performance of composite biological materials such as nacre and bone…" [26]*

The structural design principle is the usage of sacrificial connective proteins for toughness. Energy is expended to unravel these sacrificial modules, which increases the material toughness. A materials designer may be able to use sacrificial modules in the design of a new, tough bioinspired material.

## Overview of Chapters

The structural design principle algorithm research is at an intersection of design, materials, and computer science. In Chapter 2, a more in-depth discussion of each of the areas is covered. The following four chapters each describe a stage of algorithm development from research approach through evaluation. In Chapter 3, the first stage of algorithm development is examined. The initial algorithm uses information retrieval to extract passages that describe structural design principles from a materials corpus, although it was found to return many irrelevant passages as well. Chapters 4 through 6 describe the work completed towards improving the percentage of relevant returned passages. In Chapter 4, classic text classification techniques, such as supervised machine learning classifiers, statistical features, and part-of-speech analysis, are considered. In Chapters 5 and 6, two separate rounds of manual rule-based classification take place to manually identify patterns and templates that can be used to classify the passages as relevant or irrelevant. Chapter 7 validates the ability of the tool to identify structural design principles and demonstrates that bioinspired materials can be created from the design principles. Chapter 8 provides overall conclusions of the research, which includes research contributions and recommendations for future work.

# CHAPTER II

# BACKGROUND*

This chapter introduces data and text mining, and examines related research. First, it investigates search tools and databases that have been developed for general bioinspired engineering design. Next, the chapter gives an overview of data mining and discusses materials data mining, which is often used to quantitatively predict material properties or structures. The chapter concludes with a review of text mining and explores text mining work similar to this dissertation's research.

## Bioinspired Design Tools

In general bioinspired design, a designer will look to nature for inspiration to solve an engineering design problem. Currently a major issue in bioinspired design is that designers generally lack the knowledge necessary to identify biological systems relevant to their design problem [31]. However, simply increasing an engineer's biological knowledge base may not improve the probability of finding a suitable analogous natural system. A study was completed by the author comparing the abilities of senior engineering and biology undergraduates to use nature as inspiration for a

---

* Part of this chapter is reprinted with permission from "Development of a Search Tool to Identify Structural Design Principles for Bioinspired Materials Design" by Tsenn, J., Linsey, J. S., and McAdams, D. A., 2014. Proceedings of the ASME International Design Engineering Technical Conferences & Computers and Information in Engineering Conference, DETC2014-34291. Copyright 2014 by ASME.

simple concept generation problem [32]. The students well-versed in biology were not able to identify significantly more analogous systems or develop more creative solutions. The results suggest that bioinspired design methods or tools may be needed to give a more organized approach to finding analogous solutions. Glier et al. examined the directed intuitive method of bioinspired design, which simply directs the designers to consider how nature may solve a problem [33]. The authors discovered that the directed approach offered no advantage in terms of quantity, quality, novelty or variety of the solutions produced by novice engineers, and recommended the usage of more formalized and systematic methods and tools to effectively leverage nature's design knowledge.

Search tools and databases have been developed to overcome the main issues with bioinspired design. As mentioned, the first issue is the designer's lack of deep biological knowledge. Second is the terminology differences between the biology and engineering domains, which make it difficult for a designer to locate and understand pertinent information. The third issue is the design engineer's difficulty with identifying which biological solutions are analogous to their engineering problems. A majority of the engineering bioinspired design tools make use of the design problem's function, such as transferring a load, storing energy, or coupling members, to search for and categorize relevant biological knowledge and overcome the terminology differences [34-40].

One approach to identify biological information relevant to the engineering design problem is the usage of systematic search tools. Designers can use the created tools to locate pertinent information from previously generated biological text. The advantage of using a search tool is that a large amount of biological information exists in the natural language format, and the user can take advantage of the journals and textbooks available. Search tools are also adaptable and scalable as the corpus text can be changed. However, a disadvantage to the utilization of search tools is that the tools may return a large amount of irrelevant information. Search tools can address the issue of different domain lexicons used by engineers and biologists. By searching for functional keywords in the form of verbs, engineers can extract pertinent biological phenomena from biological sources. At the University of Toronto, methods were developed to allow natural-language based searches within biological text by algorithmically generating relevant biological keywords [39, 40]. In a separate effort to bridge the gap between engineering and biology terminology, Nagel et al. developed an engineering-to-biology thesaurus [36]. Terms from the engineering Functional Basis developed by Hirtz et al. [41] are paired with biological terms, providing biological search terms to an engineer. BIOscrabble was created to support the search for biological analogies with a different search term formulation [42]. WordNet, an electronic lexical English database, can also be used to identify alternative keywords to assist with cross-domain searches [40, 43]. Vandevenne et al. developed methods and algorithms for a systematic search identifying relevant biological information with an emphasis on scalability [44-46].

The other principal approach to finding applicable biological information is through the usage of developed databases. The databases are organized by categories familiar to engineers, often functions, to help engineers identify applicable biological information. An advantage of utilizing databases is that the information within the database is generally highly relevant and useful. The disadvantages are that compiling and updating the databases are resource intensive, and the data is subjected to compiler bias. Chakrabarti et al. developed the Idea-Inspire software to search two of their databases to find pertinent entries [34]. One database consists of approximately 100 entries of natural systems from the plant and animal domain, while the other database contains entries of mechanical systems. The entries use the causal description language SAPPhIRE to describe functionality and behavior. A separate search tool known as DANE was developed to help identify biological phenomena in order to assist in solving engineering problems [38]. DANE uses Structure-Behavior-Function (SBF) modeling to categorize biological and engineering systems, and its interactive design library allows for a search of the models by function. Biologue also makes use of SBF models for biological and engineering systems, but is a social citation cataloguing system [35]. AskNature is an open source project that organizes biological literature by function to support bioinspired design [37]. Finally, BioTRIZ is essentially a database of biological system solution principles [27]. BioTRIZ was developed following the study of approximately 500 biological phenomena and 2,500 biological conflicts and their resolutions in order to develop a matrix derived from the Altshuller's standard TRIZ matrix [47].

This dissertation's research involves the development of a bioinspired search tool, although it is geared towards materials design rather than general engineering design. Similar to the general bioinspired design tools, one of the major goals is to overcome the knowledge gap between materials designers and biological knowledge by searching through a corpus. However, there are key differences between this dissertation's work and the other presented research. One difference is the search terms that are used. Because materials designers are generally more interested in identifying materials with specific properties, as opposed to concentrating on function, the main search terms are property terms. Other search term sets are required because the dissertation tool needs to mine for structural design principles. Rather than a simple search of the property terms, the algorithm is developed to identify relationships between material properties and structure. Another difference is that the lexical gap between engineers and biologists is not as significant in the materials domain. The program primarily searches materials publications, and the materials domain has its own terminology, which is congruous for both biological and engineered materials. Since the expected end user of the program is a materials designer, they should be familiar with this domain language. Although there may be some biological terminology unfamiliar to the materials designer, such as specific mineral forms or scientific names, a majority of the text discusses material properties, structures, and testing, which a materials designer can comprehend.

## Data Mining

Data mining involves the extraction of patterns and discovery of information from data [3]. When the data set is very large, it is advantageous to use computer-based methods to determine relationships, patterns, and models of system variables rather than using analysts. There are two primary goals of data mining: prediction and description.

Predictive data mining creates a system model that can be used to predict new values based on the input variables.  Descriptive data mining examines the data to generate relationships and patterns that can then be interpreted by humans.  There are also two types of learning methods: supervised learning and unsupervised learning.  In supervised learning, an external teacher or method exists that can estimate the relationship between the input and output data.  For example, a human might label data prior to using a supervised learning algorithm.  The algorithm then analyzes this training data to produce a function that can apply labels to new data.  For unsupervised learning, input values are given to the system, and an unsupervised learning algorithm attempts to find the inherent patterns without the use of an external teacher or method.

The data can be classified into three categories, which can affect the type of data approach to take.  Data may be *structured* with well-defined numeric or alphanumeric values, as would be found in a data table.  *Semi-structured* data has a less constrained schema and includes text files or web pages.  Video or other multimedia recordings would fall within the *unstructured* data classification.  Semi-structured and unstructured data require more processing in order to extract the valuable information and impose a structure for analysis.

There are six primary tasks of data mining [3, 48].  Classification is a supervised function that classifies items into target categories.  Common examples of classification include email spam identification and sentiment analysis for grouping reviews into positive and negative classes.  Regression is a supervised function that predicts a real value number for a data item.  Clustering, an unsupervised task, is used to group data items that are similar to each other into clusters with low inter-cluster similarity and high intra-cluster similarity. Summarization is a descriptive task that provides a compact description of the data. Association rule learning is an unsupervised task that searches for association and co-occurrence between variables. One usage of association rule learning is to determine items that are frequently bought together in a single transaction, which can then be used for marketing. The sixth primary task is anomaly detection, which identifies unusual data in the records. Anomaly detection may be used for detecting fraud.

Text mining is a special field within data mining that extracts nontrivial information and discovers knowledge from large amounts of semi-structured textual data. Text mining investigates relationships both within and between documents for a particular purpose [49]. Due to the ambiguity that is inherent with semi-structured text, text mining has been recognized to be more complex than traditional data mining [3].

## Materials Data Mining

Materials design has historically been completed on a trial-and-error basis with the designer using his own experience to predict the different combinations of components or processing needed to attain the desired properties [28]. Data mining has recently become more popular because it can more accurately find correlations and make predictions with the use of available data. Another push towards data mining for materials design is due to the fact that experimental and simulation dataset acquisition has become more automated. The amount of data collected is outpacing the amount of analysis that can be completed. There has thus been a move towards a data-driven effort to process the experimental and simulation datasets for a more systematic materials design process [50]. Predictive data mining is used in materials design to quantitatively characterize materials from structured datasets through statistical correlation determination.

In Morgan and Ceder's book chapter, they introduced the concept of using data mining for materials analysis, describing common methods and applications in materials development which could be accomplished through data mining techniques [28]. They reviewed the prediction of physiochemical properties using a database of structures and properties, as well as the modeling of electrical and mechanical properties, and the prediction of crystal structures using a database. Data mining was used in 2001 by Chalk, Beck, and Clark to predict a material's boiling point [51]. They used a database of molecular structures and boiling points to determine a set of 18 variables and were able to develop accurate predictions for a test set of materials. Ortiz, Eriksson, and Klintenberg developed a method to predict electronic structure using an inorganic crystal structure database, and were even able to identify new functional materials [7]. Tuchbreiter and Mulhaupt describe a high-throughput process, which allowed for a quick creation of a database using the results of many polyolefin experiments [52]. A software program then quickly evaluated the data for correlations between the material properties and spectroscopic data. Rajan surveyed this form of combinatorial materials science experimental strategies and discussed high-throughput screening to create data libraries [53]. Materials data mining can even utilize fundamental quantum mechanics calculations to predict crystal structures and properties by mining experimental observations [54, 55]. Other forms of materials data mining research are more related to the advancement of the data mining process. For example, Stephenson et al.'s paper discussed the development of a new clustering algorithm, which could be used to characterize atomic clustering in atom probe tomography [56]. With all of the useful materials knowledge available, research into how to create material databases for relational statistical analyses has been completed. One such database includes an environment for the quantitative comparison of microstructures [50]. A material selection database, with a compilation of quantitative material properties, has been used to begin developing relationships within different classes of materials [57].

Currently, materials data mining is primarily used to predict numerical properties, specific crystal structures, or other material measures. The goal of this dissertation work is more descriptive in nature since it is identifying potential trends and relationships, rather than quantitative predictions and correlations. While most of the other materials data mining programs use quantitative structured property databases and experimental results, this dissertation research uses semi-structured data from published journal articles or textbooks. Descriptive data mining is required to identify passages containing structure-property relationships. A materials designer is needed to read through the passages and identify the structural design principles that can be used for bioinspired materials. Rather than predicting quantitative characteristics for a particular engineered material, the end product is the discovery of material structures and general trends that may be used with an assortment of components to produce a variety of bioinspired materials.

## Text Mining

The goal of text mining is to aid researchers in identifying useful information efficiently from the large amount of available textual data [49]. One purpose of automatic text analysis is to provide an overview of the information within the text and efficiently organize the information for the user [3]. The analysis can also be used to improve upon the effectiveness and efficiency of a search process to identify relevant information. A third purpose for text analysis is to extract hidden structures both within and between documents.

Text mining can be achieved through a wide variety of tasks. Information retrieval is used to locate useful information from a large number of documents, such as a web search. Often, the retrieved information is in the form of full documents and the documents are rank-ordered in terms of their relevance to the query [49]. Question answering is also within the information retrieval domain, and is further divided into research areas including query expansion, result ranking, and answer extraction. Lin et al. showed that users prefer passages as a response to question answering systems over both the exact answer and a full document because the exact answer and full document provide too little and too much information, respectively [58]. Textual analysis can also follow traditional data mining for the classification, clustering, or summarization of documents. Another area of text mining is information extraction to obtain structured data from unstructured text. Information extraction transforms documents into a format ready to enter into a database and includes named entity recognition (NER) and relationship extraction. NER is used to identify all occurrences of a pre-specified type of element within text, like names of people or locations [49]. Relationship extraction builds on NER by searching for types of relationships between known named entities.

For information retrieval, the two basic metrics most frequently used to measure system effectiveness are recall and precision [59]. Recall is defined as the fraction of the relevant documents that are successfully retrieved by the algorithm (Equation 1). Precision is the fraction of the documents retrieved by the algorithm that are relevant (Equation 2). Recall and precision can also be used to evaluate text classification algorithms based on correct or incorrect document classification.

$$Recall = \frac{\#(relevant\ items\ retrieved)}{\#(relevant\ items)} = P(retrieved|relevant) \tag{1}$$

$$Precision = \frac{\#(relevant\ items\ retrieved)}{\#(retrieved\ items)} = P(relevant|retrieved) \tag{2}$$

## Related Text Mining

From a text mining standpoint, the most similar work to the overall structural design principle identification research is within the biomedical text mining field. Like this dissertation research, the biomedical field has an enormous amount of available information, and requires the use of text mining to uncover the relevant information efficiently [49, 60]. Researchers are creating algorithms to mine biomedical abstracts and articles for relationships between known entities such as genes, proteins, and drugs [49, 61-64]. However, there are some key differences between biomedical text mining and the biological materials text mining research. One of the major differences is that the genes, proteins, and drugs can be located and classified prior to searching for the relationship using Named Entity Recognition (NER). While the properties can be identified in the articles for this dissertation research, the material structures are unknown and are not a form of named entity. Another key difference is that the biomedical field makes use of UMLS, the Unified Medical Language System. The UMLS is used for biomedical informatics and consists of a controlled vocabulary and ontology of concepts, which makes biomedical documents more consistent [65]. There is not a similar language system within the materials domain. A difficulty shared by both research areas is the need for domain expertise for algorithm development in their respective fields.

The portion of this dissertation research attempting to reduce the number of irrelevant passages is most similar to the text mining technique of text classification. The classes would consist of the “relevant passage” class and the “irrelevant passage” class. Text classification techniques will be further discussed in Chapters 4 through 6 with their corresponding evaluation. Domain-specific text classification is difficult, and generally results in either low recall or low precision, or passable values for both due to the inverse relationship between the two metrics. An example of text classification within the biomedical field was completed by Dobrokhotov et al. They used text classification to determine if article titles and abstracts retrieved using a gene name and protein search should be curated into the Swiss-Prot database. Using a Probabilistic Latent Categorizer with Kullback-Leibler symmetric divergence, a recall of 69% was achieved with a precision of 59% [66]. Another form of biomedical text classification is sentence type identification. McKnight and Srinivasan were the first to automatically classify sentences from biomedical abstracts into introduction, method, result, or conclusion sentences [67]. Using unstructured MedLINE abstracts, they were able to correctly identify the introduction sentences using a linear classifier with a recall of 34% and a precision of 71%. By using support vector machines, the recall increased to 49% and precision to 82%. The aforementioned text classification studies benefit from using full abstracts, which include full thoughts, while this dissertation research requires the extraction of short passages and utilizes those passages for classification purposes. The biomedical domain also makes use of the UMLS, which generally aids in classification.

## Conclusion

This chapter examined bioinspired design, data mining, the usage of data mining in materials design, text mining, and related text mining work. The exploration of current bioinspired design tools demonstrated the value of search tools in identifying relevant biological analogues to designers. Next, a discussion of data mining and the associated tasks took place. The purpose of data mining in materials design was often predictive in order to design specific materials and identify quantitative properties based on other variables. An overlap between bioinspired design and data mining for the development of bioinspired materials was not found. A brief overview of text mining was also included. The research in this dissertation was found to be most similar to the biomedical text mining area due to its search for relationships and classification of text, and a few significant differences were noted. This dissertation research begins the exploration into the unique field of text mining algorithm development to aid with the design of bioinspired materials.

# CHAPTER III

# ALGORITHM DEVELOPMENT STAGE 1: INFORMATION RETRIEVAL*

For this dissertation's research, a text mining algorithm that can quickly identify potential structural design principles of biological materials with respect to chosen material properties is developed. This chapter presents the first stage of algorithm development: information retrieval. The Stage 1 algorithm is designed to extract structural design principles' key terms and relevant passages for specified material properties from a corpus of materials journal articles. The development of the search tool can be broken down into two main phases: the development of the search algorithm and the determination of the search terms. Many articles were studied to refine these general needs into specific needs and requirements. Following a discussion of the research approach, an evaluation of the algorithm is completed comparing the Stage 1 algorithm's results to those of a manual search for the structure-property relationships.

## Algorithm Development

The general search algorithm reflects how a user would interact with the program interface. The interface accommodates a user's selection of a material property or

---

* Part of this chapter is reprinted with permission from "Development of a Search Tool to Identify Structural Design Principles for Bioinspired Materials Design" by Tsenn, J., Linsey, J. S., and McAdams, D. A., 2014. Proceedings of the ASME International Design Engineering Technical Conferences & Computers and Information in Engineering Conference, DETC2014-34291. Copyright 2014 by ASME.

combination of properties. The algorithm then systematically searches the corpus for biological articles that explain how the material's structure affects those desired properties. Next, passages describing those identified structures are saved into a text file, which is returned as the program's output. Since the designer can now read through the short passages rather than entire articles, the algorithm reduces the time a materials designer needs to spend searching for biological inspiration.

After conducting a search for biological and natural materials design papers, seventeen highly cited, relevant papers were chosen to aid in the development of the more detailed search algorithm displayed in Figure 4. The papers covered topics including hierarchical biological materials, biomimetic structural materials, natural composites, and plants. While there are many biological journals and materials journals, there are few that are solely dedicated to the structure and properties of biological materials. As such, the search algorithm needs to be able to first identify if an article is indeed about a biological or bioinspired material using biological search terms. A biological or biomimetic paper generally gives an introduction to biological materials in the beginning of the paper, which is far from the passages describing the structure-property relationships. Thus, the biological search needs to take place separately from the search for the structure-property relationships. The algorithm first searches through each article individually for a biological search term. If a biological term appears, then the algorithm investigates the paper further for properties. If no biological terms appear, then the algorithm moves to the next article. The search terms used to signify a biological or bioinspired paper were carefully chosen and are described in the next section of this chapter.

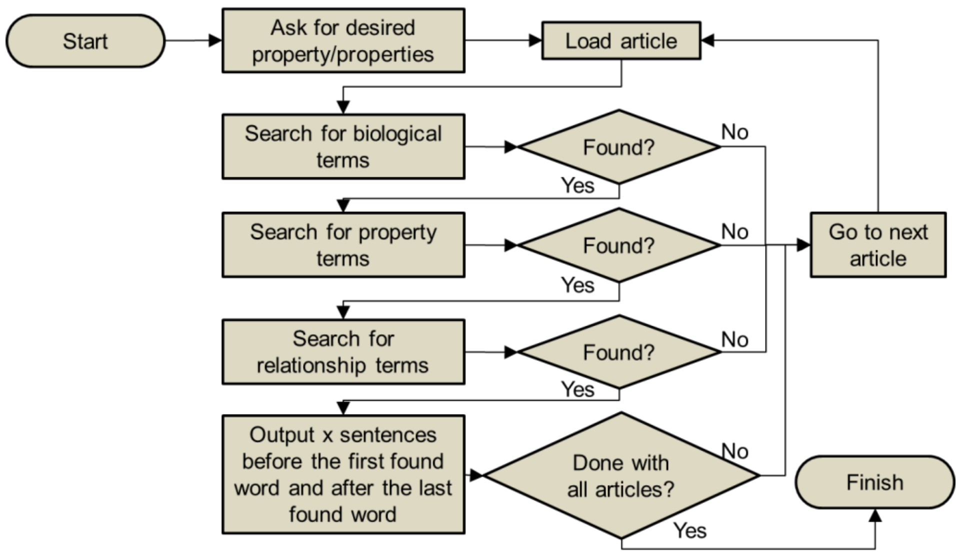



Figure 4. An outline of the Stage 1 algorithm.

The biological materials papers then typically describe the material's properties. Only the property discussions relating to the material's structure are of interest. A relationship term is used to describe the association between a material's property and its structure. To identify the passages, the algorithm searches for property terms and relationship terms. If a property term and a relationship term appear within a specified number of sentences of each other, then the passage is saved. A majority of the passages identified in the biological materials papers are one to two sentences long, although several are three to four sentences in length, with very few that are longer. To take into account these passage lengths, variable search and return ranges are added to the algorithm. The Stage 1 algorithm checks for property and relationship terms that are within two or fewer sentences. It then saves a passage that includes the two sentences prior to the first search term through the two sentences after the last search term. With the described algorithm, the passage should include the structure-property relationship, regardless of if the properties or the structures are described first.

The search algorithm was structured in a particular way in order to exclude general descriptions of properties and unrelated usage of the search terms for more meaningful results. Using the user-selected properties, the algorithm first searches through each article for the predetermined property-specific collection of search terms. If only one property was initially selected, the algorithm notes the sentences in which the property's search terms appear. After completing the search, the algorithm hunts for a set of relationship terms to identify causations and associations. The algorithm then saves passages with the set passage return length. With this structure, passages returned by the algorithm should describe the property of interest along with its relationship to the structural design principle. For the nacre example seen in Figure 2, key words describing the structural design principle would be "layered structure" and "brick-and-mortar structure".

In order to learn how nature is able to balance a combination of properties, the algorithm needs to allow a user to select multiple properties at the start of the algorithm. A few changes need to be made to the search process for this scenario. The algorithm searches for the predetermined search terms for each of the specified properties and records the sentences in which they occur. Next, the algorithm determines the number of sentences separating the different property terms. By only keeping sentences that contain terms from the different selected properties within the preset sentence range, the algorithm is able to refine the selections to only include those that describe all of the user-selected properties. After completing the property term searches, the algorithm searches for the relationship terms. Once again, only the sentences that are near a relationship term are returned as output passages.

The materials designer can search the formatted materials journal articles available in the program, but can add or substitute in additional text as well. To do so, a text file with the articles simply needs to be supplied in the appropriate folder and formatted with a double hash ("##"), the article information, and a single hash prior to the start of each article's text. The program output consists of a text file containing the supplied articles' information followed by the passages returned from each article. If the user is interested in a particular material's structure, he can then easily find the referenced paper for further study.

## Search Term Selection

There are three sets of search terms needed for the algorithm. First, a set of biological terms is needed to check if a materials paper is about a biological or bioinspired material. Secondly, a set of search terms is identified for each of the predetermined material property options. The third set of search terms is relationship terms, which are needed to identify which passages to extract. The process used to find the proper search terms for each of these categories is described in the following sections. These initial sets of search terms are used to test and further develop the algorithm. The search term lists may be updated during algorithm testing as additional search terms are found.

### *Biological Search Terms*

To determine the biological search terms, the seventeen biological and bioinspired materials papers used for algorithm development were examined for popular biology-related terms. As previously mentioned, the abstract or introduction of the paper often gives a brief explanation of biological materials or introduces the biological material directly. The words that appeared often in the introduction to describe biological materials were noted, and the most commonly used words were further studied. The most popular terms were found to be “biological”, “nature”, “natural”, “organic”, and “inorganic”.

Because the program is designed to run quickly, the terms list does not contain all of the biology-related terms that were found. Ideally, the list would only include the minimum number of biological terms that appear in every biological materials paper. Since a word frequency analysis would take time to develop, additional terms were found by studying the previously identified biological terms using Thesaurus.com [68], Wikipedia [69], and WordNet, which is an English lexical database [70]. The thesaurus and WordNet were chosen because both databases contain lists of words with similar meanings, allowing for the identification of synonyms, hypernyms, and troponyms for the commonly used biological terms. Wikipedia was used because it contains a description of the term, which likely uses similar terminology to biological articles describing biological materials. Some of the identified terms were too general (such as "outdoors"), some were too specific (such as "embryology"), and some results could have different meanings depending on the context (such as "environment"). While these words could be found in biological articles, they would either require too many search terms (especially when delving into the names for different fields of biology) or they would likely return many non-biological articles. The more common terms were chosen for the biological search term list. Then, the seventeen various biological materials papers were examined again to determine if at least one of the search terms appeared in each of the papers. Three of the papers did not contain a search term, so the most general biological term from each paper ("biopolymer", "protein", and "tissue") was added to the list. A total of 25 terms were finalized for the biological search term list, and the terms can be found in Appendix A. Because the algorithm needs to match the terms exactly, some of the words are related, such as "biopolymer" and "biopolymers".

*Property Search Terms*

The preselected property search terms are encoded into the tool. The users will only need to input the property or properties they want to study. By using only predetermined search terms, the user will not unknowingly restrict the search space with biased keywords. Also, since Hölscher and Strube found that those with novice technical query-based search skills or less domain-specific background knowledge are less successful in their search behavior [71], using set search terms is advantageous because it allows the results to be independent from the user's search abilities.

A list of potential properties for the algorithm was generated using textbooks, web searches, and Wikipedia. In order to concentrate on the program algorithm, rather than property search terms, the property options were narrowed down to six properties for the development phase of the algorithm. For now, the six material properties available are strength, ductility, stiffness, toughness, hardness, and density. These properties are important for structural materials, which is a common application for bioinspired materials. The property options will be expanded in the future.

A different collection of search terms is used depending on the property selected by the materials designer. Materials textbooks were studied to identify potential search terms for each specified property. The books were selected based on the criteria that they include descriptions of properties and use commonly related words and phrases that are likely to appear in biological materials papers. The sections of Callister's Materials Science and Engineering textbook [72] relating to material properties for metals, ceramics, and polymers were read, and all of the terms relating to each property were recorded. The same process was also completed using Ashby's Materials Selection in Mechanical Design [57] since the book introduces and discusses each of the structural properties in detail. Each property was then searched using Wikipedia, a thesaurus, and WordNet to identify additional search terms for the potential property terms lists.

Once the lists of potential search terms were compiled, each list was refined based on certain criteria. The search terms need to be able to identify a majority of the passages related to a property, while not being so vague as to return irrelevant passages. For example, if the user is interested in the strength of the material, passages including the terms “ultimate tensile strength” will likely be useful. However, specifically searching for “tension” may produce passages about an experimental procedure or may be using the non-scientific definition of the word. The thesaurus and WordNet did not yield useful property search terms because they use the more commonly used definitions of the property terms. The search terms also need to be specific to a single property. For example, using “strain” as a search term could return information for strength, ductility, stiffness, or toughness. If the selected property is ductility, the entire phrase “strain to failure” would be a better search term. However, if a search term accurately encompasses multiple properties, then the term is kept, such as “specific bending stiffness” referring to the properties of stiffness and density. Search terms were also removed if they were so specific that they were not likely to be used as descriptor phrases in biological materials papers such as the “resistance to separation of adjacent atoms”, used to define the modulus of elasticity in Callister’s textbook [72]. A total of 67 of the 104 (39%) identified terms were removed for being too vague to refer to a specific property or too specific to likely be used in future descriptions.

Many of the search terms were written in multiple forms. For example, “stiff” was also used as “stiffness” or “stiffen”. To identify all of the terms relating to “stiff”, one option would be to stem all of the search terms to return all related passages, a process known as query expansion. However, this reduces precision of the returned results and would cause the algorithm to run more slowly. While a query expansion would be useful for a web search or user-entered search terms, precision is more important for the design principle algorithm and can be taken into account with the pre-set terms. In order to attain more accurate results, many of the search terms are actually phrases. For example, “yield strength” is used as a strength property search term and “impact strength” is used as a toughness term. By using these phrases rather than the general term “strength”, many extraneous results can be eliminated.

Since the search needs to be very specific, the exact word or phrase needs to be found in the corpus. Therefore, different morphological versions of the terms are included in the property search terms. The plural version of nouns, different tenses of verbs, various comparative forms, and even the reorganization of words in a phrase (“resist fracture” and “fracture resistance”) are in the search term lists. A total of 20 terms were chosen for strength, 14 terms for ductility, 11 for stiffness, 25 for toughness, 28 for hardness, and 6 for density (Appendix A).

### *Relationship Search Terms*

A relationship term is needed to find passages that indicate causality or association. These search terms will help locate the passage in which the relationship between the material’s structure and its properties are described. The relationship terms were determined by reading through the seventeen biological material articles and searching for relationship descriptions. The terms include words and phrases like “consequently”, “due to”, “provides”, and “relating”. As more papers were investigated, the number of new terms added to the relationship search term list decreased. By the seventeenth paper, no new terms were added to the list of 44 unique terms. To account for all morphological variations of these terms, there are a total of 141 relationship search terms (Appendix A). While more terms could be found using a thesaurus or WordNet, adding all of the possible terms would make the list too long and cause the algorithm to run for an unnecessarily prolonged period of time. Since the key structures and properties are repeated multiple times throughout the paper, having a list that includes most of the relationship terms should be sufficient to identify the principles.

### *Excluding Functional Terms*

The decision was made to not include functional terms in the biological materials search algorithm because function is often directly related to property. For example, when a designer is interested in the property of “adhesion”, they are also interested in the function of “couple”, as described in the Functional Basis [41]. Other properties similar to those provided by the Functional Basis include sensing and actuation. By searching for property terms and excluding functional terms, the program can run more quickly. The property terms are better suited for the algorithm because they are a part of the materials science domain language, and they are able to represent material traits more precisely than functional terms. For example, the properties of “strength” and “stiffness” are distinct, but are both part of a structural materials’ main function of supporting a load.

## Research Approach

The algorithm was tested by completing a manual search for structure-property relationships in a journal issue and by running the same journal issue through the search program. This section further details the approach used for the comparison. The section also examines the search terms used in the journal issue and describes the refinements made to the search term lists. Finally, a comparison of the manual search's results and the algorithm's returned passages is investigated. The passages output by the algorithm are inspected for the manually identified key structural design principle terms describing the structure-property relationships.

### *Comparison Approach*

To determine the refinements needed for both the algorithm and the search terms, an analysis of one journal issue was completed. A single issue was selected because it can provide a variety of topics, viewpoints, and writing styles, while still having a central theme. Advanced Materials Volume 21, Issue 4 [73] was chosen because it is a Biological and Biomimetic special issue containing 19 peer-reviewed submissions. Half of the articles in this issue discuss biological materials while the other half relate to bioinspired materials. The comparison strategy for the results is discussed in this section, and the results themselves will be presented in the following section.

A manual search for principles describing structure-property relationships is completed for each article in the issue. The goal of the algorithm is to identify how material structure and arrangement affect specific material properties, thus the manual search ignores passages that vaguely describe the effects of the structure such as “it improves mechanical properties”. After finding all relevant principles, the descriptive property terms and their corresponding mechanical properties are identified. For example, a descriptive term found in a principle passage may be “rigidity”, which corresponds to the property of stiffness. Then the relationship term within each passage is manually found. If any of the property and relationship terms do not appear in current search term lists, they are noted. The results of the manual search are entered into a matrix to relate each property and combination of properties with the key terms describing the structural design principle.

Following the manual search, the algorithm examines the same nineteen articles using the initial search term lists. Each of the six properties is run individually along with all possible combinations of the properties for a total of 63 runs. The results from the algorithm are compared to the manually produced matrix to determine if the expected principle key terms are present for each property combination. A variation on the well-known information retrieval metric recall is used. For this research, recall is defined as the percentage of the manually identified principles that are successfully retrieved by the algorithm. If the recall value is low and changes need to be made to the biological, property, or relationship search term lists, the program will be rerun for all the properties again with the updated lists.

If the algorithm is still not returning passages containing the structural design principle key terms, the algorithm’s variable search and return ranges can be modified. If the sentence range is combining too few or too many properties, the maximum number of sentences between the property terms can be updated. If the algorithm captures the structure and property's relationship along with extraneous, unrelated sentences, then the number of sentences between the property and relationship term can be reduced. If the passages are not long enough or are too long, then the length of the returned passage can be changed as well. Finally, if modifications are needed for the search terms or the algorithm structure, then the comparison process will be repeated for a new journal issue with all new articles.

### *Manual Search Terms Results*

Each article was manually investigated to see if it contained at least one of the previously determined biological search terms. It was found that the biological search term list is comprehensive enough to encompass all of the articles in the Advanced Materials Volume 21, Issue 4 Biological and Biomimetic special issue [73], so no changes were made to the biological search term list. Since all of the articles passed the first check, each was subjected to a search for structure-property relationships.

For the manual search, passages describing a structure and its effects on specific mechanical properties were identified for each article. From each passage, the property search terms, relationship search terms, and key structural design principle terms were ascertained. The passage referenced in Chapter 1 was selected to demonstrate this process:

> *"…while tightly packed domains tend to resist unfolding, AFM-based single molecule pulling experiments on the muscle protein titin revealed that unraveling of loosely packed modules led to increased molecular contour length and an increase in the energy absorbed prior to*

> *catastrophic failure. Similar sacrificial connective proteins have been proposed as toughening agents in the macroscopic mechanical performance of composite biological materials such as nacre and bone…" [26]*

From this manually identified passage, the property search terms are "energy absorbed" and "toughening", which both fall within the property of "toughness". The relationship search terms include "revealed" and "led". The identified principle is the relationship between the use of sacrificial connective proteins and the increase in toughness due to the energy expended to unravel the modules. The key principle terms are then "sacrificial connective proteins".

For the property search terms, it was discovered that there are a variety of additional terms that are used to indicate a particular property. For example, stiffness is not described just by "Young's modulus" or "stiffer" or "rigidity", but can also be described as "resistant to bending" or "compliant". Phrases referring to strength could be "theoretical strength" or "load-bearing" or even "high loads". A second set of search terms were created to take into account these additional terms. However, only those search terms considered to be valuable were added; this included all but two of the new terms. Phrases that are uncommon variations of more popular phrases were not added. For example, Porter and Vollrath described the effects of hydration converting fibrous proteins from a hard state to a rubber-like state as "capable of large strains and absorbing large amounts of mechanical energy" [74]. The phrase "absorbing large amounts of mechanical energy" was not added to the toughness property search terms because "large amounts of" is not a part of a commonly used property phrase. A total of 13 new search terms were added to the updated “strength” property list, 8 to “ductility”, 8 to “stiffness”, 10 to “toughness”, 2 to “hardness”, and 7 to “density” (Appendix A).

Through the manual search, many potential non-structural properties were discovered. One of these popular properties is adhesion, the joining of two materials. Since this is a property prevalent in nature and relevant to many engineering design problems, it would be beneficial to add adhesion to the property options for the algorithm in the future. Other properties to consider include reactivity, wear resistance, and self-cleaning. The functions of sensing and actuation are also common in nature and would be invaluable for the design of bioinspired materials.

As with the property search terms, there were a variety of ways in which the relationship search terms could be worded. Many of the words and phrases were repeated from the previously determined search term list, although there were still a number of terms found which were added to the revised list. These new terms included expressions like “attributed to” and “implications”. There were a total of 18 different new terms identified, and 86 terms were added to the revised search term list in order to take into account all of the morphological variations of those 18 new phrases (Appendix A).

As more articles are studied, new property and relationship terms are added to their respective revised lists. However, as articles continue to be read, the number of new terms should decrease and plateau asymptotically. Although there are many possible terms and expressions, the algorithm does not need all of them in its search term lists, as a technical paper will often repeat important concepts multiple times with different wordings for emphasis and clarification. The structural design principle key terms may appear in multiple sections of a paper and the algorithm merely needs to identify them once. When searching a large number of articles, it is likely that a structural design principle will show up in multiple articles, leading to the key terms appearing a variety of times.

### *Comparing the Manual and Algorithm Search Results*

An investigation of the algorithm-extracted passages for the manually found structural design principle key terms took place. The key principle terms are the important phrases that describe the material's structure. For example, instead of the full structure description: "One prominent design principle found in the hierarchical structure of such biological fibrous composite materials is the twisted plywood structure. In the lobster cuticle, it is formed by superimposing and gradually rotating planes of parallel aligned chitin-protein fibers" [1], the structure's key principle terms would be "twisted plywood structure". The reasoning behind searching for key terms is that the structure is not fully described every time the properties are discussed. A materials designer should be able to either recognize the structure by the principle terms or easily look it up.

The user interacts with the program in two ways. First, the user can add or remove text files from the corpus, allowing him or her to choose which publications to search. Second, the user also specifies the property or combinations of properties that the program should search for. An image of the user property selection interface can be seen in Figure 5. In this case, the program user input “123”, which is the search for a combination of strength, ductility, and stiffness in a single material structure.

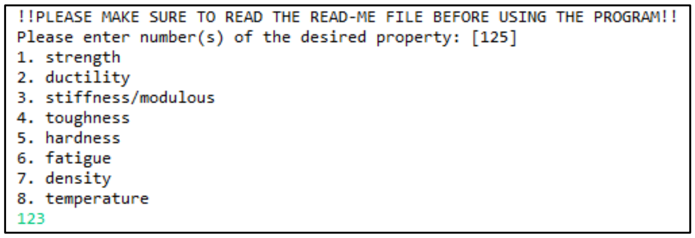
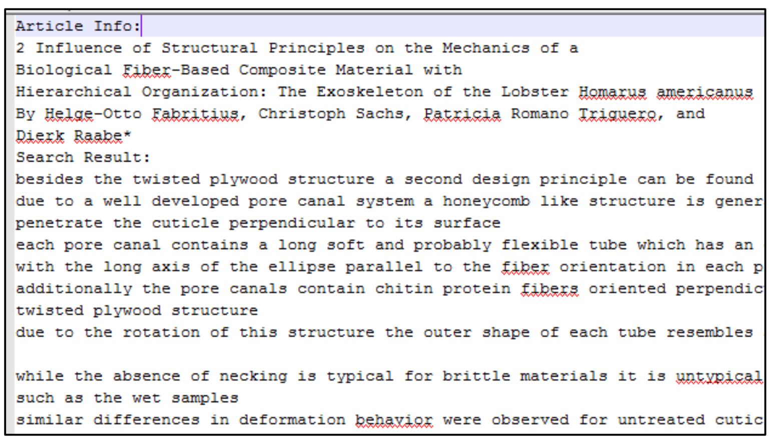

```
!!PLEASE MAKE SURE TO READ THE READ-ME FILE BEFORE USING THE PROGRAM!!
Please enter number(s) of the desired property: [125]
1. strength
2. ductility
3. stiffness/modulous
4. toughness
5. hardness
6. fatigue
7. density
8. temperature
123
```

Figure 5. A screenshot of the program’s property selection request.

To search for terms, the program makes each sentence its own line and removes all capitalizations, hyphenations, and special characters. This allows for a fewer number of search terms since various combinations of capital letters and special characters no longer need to be taken into account. However, it does affect the program’s output, as seen in Figure 6. Since there are no capitalizations, each sentence is started on a new line, and each passage found by the program is separated by a blank line. To differentiate between articles, two blank lines are inserted. The output text file contains each article’s identifying information as well as the passages that satisfy the program’s criteria. A separate text file is output for each selection of properties.

The program underwent several iterations to address a few small issues. For example, dashes were initially deleted by the program, but actually needed to be replaced with a space so that “load-bearing” becomes “load bearing” rather than "loadbearing". This was also the case for other special characters such as quotation marks, apostrophes, and forward slashes. A new rule imposed on the passages is that each sentence needs to consist of three or more words. Originally "Fig." or "Figure 1." was counted as its own sentence, but that essentially reduced the length of the output passage by one actual sentence. The program was also amended to allow more than one property search term to be found in a single sentence.



```
Article Info:
2 Influence of Structural Principles on the Mechanics of a
Biological Fiber-Based Composite Material with
Hierarchical Organization: The Exoskeleton of the Lobster Homarus americanus
By Helge-Otto Fabritius, Christoph Sachs, Patricia Romano Triguero, and
Dierk Raabe*
Search Result:
besides the twisted plywood structure a second design principle can be found
due to a well developed pore canal system a honeycomb like structure is gener
penetrate the cuticle perpendicular to its surface
each pore canal contains a long soft and probably flexible tube which has an
with the long axis of the ellipse parallel to the fiber orientation in each p
additionally the pore canals contain chitin protein fibers oriented perpendic
twisted plywood structure
due to the rotation of this structure the outer shape of each tube resembles

while the absence of necking is typical for brittle materials it is untypical
such as the wet samples
similar differences in deformation behavior were observed for untreated cutic
```

Figure 6. A screenshot of the program's output file.

For the first search, the original property terms and the initial relationship term set. There were a total of 95 principles identified by the manual results for all of the properties and their combinations in the Advanced Materials Biological and Biomimetic special issue, as seen in Figure 7. The key principle terms were then manually found and matched with its corresponding property or property combination. A search for the key principle terms was completed for each of the algorithm's corresponding output property text files. A total of 53 of the 95 expected principles were found, equivalent to a recall of approximately 56%. Because the percentage was low, the search term lists were revised using the manually found search terms in the Advanced Materials special issue, as mentioned in the previous section.

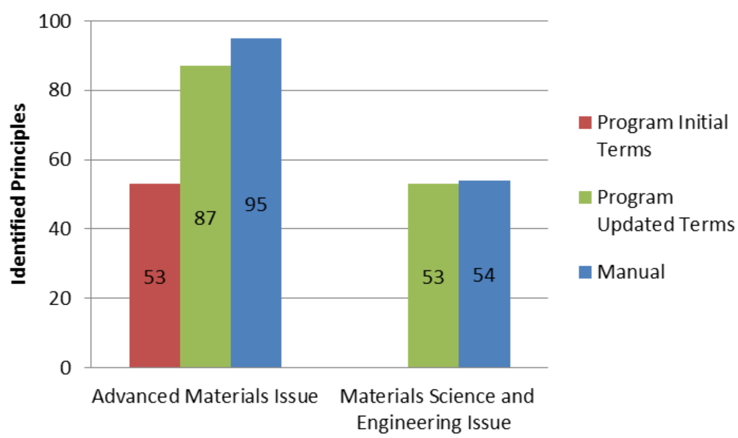



Figure 7. A comparison of the number of principles identified manually by the Stage 1 algorithm using the initial search terms and using the updated search terms.

For the second program search, which used the updated property and relationship terms, 87 of the 95 principles (approximately 92% recall) were found (Figure 7). Eight of the expected structure-property relationships were not identified by the algorithm for two reasons. First, some of the manually identified property terms were not included as search terms since the phrase was an infrequent variation of a more popular search phrase. The other reason principles were missing was because the structural key terms and the property search terms were more than three sentences apart. The few times this issue occurred, the terms were about six sentences apart. The number of sentences included in a returned passage can be increased. However, a majority of the passages do not require a longer sentence range, and changing the range would make the other passages unnecessarily long.

A final note is that there are passages unexpectedly returned by the algorithm, but these have not yet been analyzed. The metric used to calculate the percentage of relevant passages, or passages containing a structural design principle, returned by the algorithm is precision. With the updated set of property and relationship terms, the precision of the Advanced Materials special issue was 33% (Figure 8). Although there is a low precision due to the high number of unanticipated passages, it is still faster to read through the algorithm's returned results than to read through full articles. Future stages of the research will include filtering the returned passages to remove irrelevant ones so that the final tool will be able to significantly reduce the time a materials designer needs to spend actively identifying structural design principles.

## Algorithm Evaluation

An assessment of the Stage 1 algorithm's ability is completed with an entirely new set of articles using the updated search terms. An evaluation of the manually-identified principle key terms is then presented to demonstrate the repeatability of the manual results with an independent investigator.

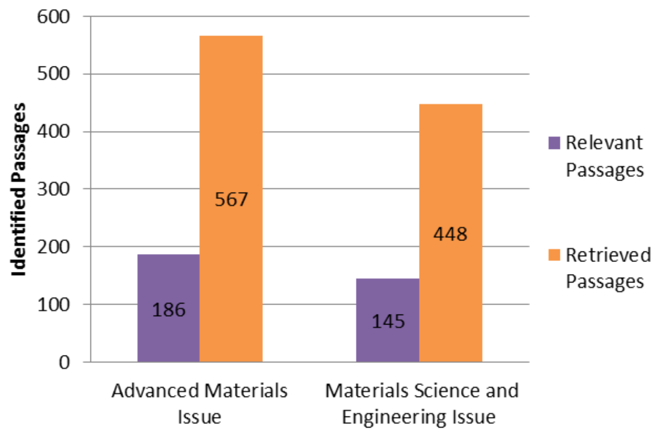



Figure 8. A comparison of the number of relevant passages and total number of passages retrieved by the Stage 1 algorithm using the updated set of search terms.

### *Algorithm Assessment Using New Articles*

A new set of articles was identified to assess the search structure and the updated search terms. The 16 articles in Materials Science and Engineering C 31 Volume 31, Issue 4 were used [75]. These articles were presented in the 6th Symposium on Biological Materials Science at the 139th Annual Meeting & Exhibition of The Minerals, Metals & Materials Society.

All of the articles contained at least one of the biological search terms. Consequently, all of the articles were searched for structure-property relationships. From the manual search, a total of 54 structural design principles were found. Of the 54 manually identified relationships, 53 were found among the algorithm-extracted passages using the updated search terms, shown in Figure 7. The only relationship that was not located by the algorithm was because the material’s structure was related to multiple properties, which were described several pages apart. The high recall of 98%

of the manually found relationships identified by the algorithm suggests that the revised property and relationship search terms are sufficient. Of the 448 passages retrieved by the algorithm, 145 were considered relevant (Figure 8), which is equivalent to a precision of 32%.

It was observed that key principle terms were already beginning to repeat between the articles in the two journal issues, which lends support to the hypothesis of underlying structural design principles in biological materials. The repeated key terms in the two issues are porous materials, sandwich structures, gradation, mineral fraction, and hydration of the material. These potential principles were identified from only 35 articles. With more articles, the principles should continue to be repeated, and new principles should be identified. Several full materials journals are currently being formatted for the algorithm's future use.

### *Evaluation of the Manually Identified Key Principles*

The manual search was completed by a graduate mechanical engineering design student that has taken nine materials courses. To evaluate the repeatability and rationality of the expected key principles found by the graduate student, a senior-level undergraduate mechanical engineering student performed the same manual search for five of the Advanced Materials articles. Of the relationships found by the graduate student, only one was overlooked by the undergraduate. Any differences in the identified properties or key principle terms were reexamined. A few of the undergraduate's errors included mixing up the properties of stiffness, toughness, and ductility based on the property terms, which would be rectified with more materials experience. A couple of the relationships listed by the undergraduate were determined to be part of other relationships. For example, the undergraduate identified the effect of the relative content of specific copolymers as a relationship, but this content was actually discussed as part of a graded materials structure.

The undergraduate did identify one of the relationships missed by the graduate student. He also identified a structure involving the “assembly of interlocking components”. While the graduate student had initially identified that relationship, she had disregarded it as being too macroscale. However, after reconsideration, the structure was added to the manual search results. Both changes were made prior to the comparison of the algorithm results to the manual results. Both relationships were found by the algorithm using the revised search terms.

## Conclusion

Structure-property relationships of biological materials need to be abstracted for use in the design of bioinspired materials. However, the manual process is time-consuming, and a better method of determining relevant materials’ structures is needed. The development of the first stage of a text mining algorithm and tool for bioinspired materials design was described in this chapter. The algorithm searches through a corpus and outputs key terms and passages describing biological structural design principles.

A materials designer can utilize the algorithm to aid in identifying structure-property relationships of biological materials. The algorithm allows a materials designer to select a desired material property or a combination of properties to study. At this time, the algorithm accounts for strength, ductility, stiffness, toughness, hardness, and density. The user may also modify the program corpus by adding or removing publications in the form of text files. The algorithm will quickly search through the articles for biological or bioinspired materials, and will then determine the relationships between the selected properties and their corresponding material component structures. Passages, which contain key terms describing nature's structural design principles, are returned in a text file as the output of the program. Materials designers can then easily read through the passages to familiarize themselves with the identified structural design principles.

The chapter discusses in detail the development of the program algorithm, the creation of search term categories (biological, property, and relationship), and the process of determining the search terms in each category. A study using a Biological and Biomimetic special issue of the Advanced Materials journal compares the algorithm results to manually found structure-property relationships from the same source. Approximately 56% of the manually identified principles were also found by the algorithm, so updates were made to the search terms and refinements were made to the algorithm. The search algorithm was then tested on a new set of articles using an issue of Materials Science and Engineering C, which resulted in a 98% algorithm success rate for identifying the manually found relationships. Although the recall of the algorithm was high, the precision was low since many irrelevant passages were returned by the algorithm as well. The next step is to improve precision to make it easier for designers to identify the principles.  With the development of an algorithm to search for structural design principles in biological materials, materials designers can more easily look to nature for inspiration.  And with the use of these structural design principles, the development of superior bioinspired engineered materials or application-tailored materials can take place.

# CHAPTER IV

## ALGORITHM DEVELOPMENT STAGE 2: TEXT CLASSIFICATION

In the previous chapter, an initial version of a text mining algorithm and tool were successfully created to quickly retrieve passages describing biological materials structural design principles from a corpus of materials journals. Although the Stage 1 tool identified over 90% of the principles (recall), many irrelevant passages were returned as well with approximately 32% of the passages being useful (precision). An improvement in precision is desired to lessen the effort required by a materials designer in identifying structural design principles. Improvements in precision correspond to decreases in recall as the two are inversely related. With the goal of reducing the end user's workload, the emphasis is shifted towards precision and returning more useful passages over identifying all possible principles. However, a variety of principles should still be retrieved so recall cannot be discounted altogether.

This chapter examines a variety of text classification techniques in an attempt to improve upon the Stage 1 program by reducing the number of extraneous passages. Text classification can be used to categorize documents into predetermined classes. The documents to be classified are the passages returned using the Stage 1 algorithm, and the class options are the "relevant passage" class and the "irrelevant passage" class. The "relevant passages" contain the biological material's structural design principle, while the "irrelevant passages" do not. The goal is to accurately classify the passages, so that those in the "irrelevant passage" class can be removed from the document set, increasing the algorithm precision.  During Stage 2 of algorithm development, the text classification techniques of machine learning classifiers, statistical features, and part-of-speech analyses, are evaluated for effectiveness in sorting passages into relevant and irrelevant classes.

### Machine Learning Classifier Study

Supervised machine learning methods are commonly used in text classification [60].  Machine learning algorithms automatically learn class patterns, and use the patterns to classify new documents.  Well-known machine learning classifier models include Naïve Bayes, logistic regression, and support vector machines (SVMs).

Supervised machine learning classifiers require two sets of manually classified documents: training documents and testing documents.  For the study, the documents are comprised of passages returned from the Stage 1 tool.   The training documents consist of a set of passages along with their corresponding "relevant" or "irrelevant" manually determined designation.  To train the classifier, the features of the training documents and the corresponding document classifications are analyzed by the classifier to determine scores and rules for the features.  For text classification, features are most commonly the document terms [59].  In the previous sentence, there are ten document terms, which include "for", "text", and "classification".  The testing documents are then run through the trained classifier.  The trained classifier scores the testing documents based on their features to determine each document's class.  The classifier is evaluated by comparing the manual classification of the testing documents to the supervised machine learning classifier's classification.

Naïve Bayes is a simple linear probabilistic classifier that is fast, easy, and highly scalable. The classifier is trained by computing probability scores for the terms in each class. The generative model attempts to maximize the joint likelihood of both data and classes using relative frequency as weights [76]. The drawbacks to Naïve Bayes classifiers are that (1) they require the assumption of no interactions between features and (2) they have been shown to be outperformed by logistic regression, linear regression, and SVMs [77]. The logistic regression models and SVMs are discriminative models. Logistic regression approaches are probabilistic and seek to maximize the conditional likelihood of classes. Benefits of logistic regression over Naïve Bayes include the ability to take into account feature dependencies, the ability to add more training data in the future, and a high accuracy performance. However, logistic regression classifiers tend to not perform well with a large number of features [78]. SVM classifiers create a hyperplane in a multidimensional space to delineate between different classes. SVMs can handle non-linear feature interactions and large feature spaces, but are memory-intensive and not as efficient.

### *Machine Learning Classifier Research Approach*

In this study, the training document passages were selected from the Journal of the Mechanical Behavior of Biomedical Materials Volume 4, Issue 5, which is a Special Issue on Natural Materials / Papers from the Third International Conference on the

Mechanics of Biomaterials and Tissues [79]. The new set of articles were run through the Stage 1 search tool, and the resulting passages were given a gold standard manual classification as belonging to the “relevant passage” class or “irrelevant passage” class. The training documents were a subset of the returned passages, and included 120 relevant passages and 379 irrelevant passages. The passages were randomly selected using a sequence generator [69] so that the training documents did not all share the same author or describe the same property. Each passage in the training file was labeled using the gold standard classification with a 1 if the passage belongs to the “relevant passage” class and a 0 for the “irrelevant passage” class.

Two sets of testing documents were used. The first set contained the remaining 43 relevant passages and 100 irrelevant passages from the Mechanical Behavior of Biomedical Materials journal issue. The second set of passages were a randomized set of 49 passages from all three of the tested journal issues, consisting of the Mechanical Behavior of Biomedical Materials issue, the Advanced Materials journal issue [73], and the Materials Science and Engineering issue [80]. This testing set had 23 relevant passages and 26 irrelevant passages.

The Stanford Natural Language Processing Group’s classifier was utilized [81, 82]. The Naïve Bayes generative classifier was initially chosen for this study because it is computationally efficient and easy to use. A probabilistic maximum entropy classifier, which is equivalent to a multiclass logistic regression model, was also tested.

*Machine Learning Classifier Evaluation and Discussion*

An example of the program output can be found in Figure 9. The output includes the passage, followed by the gold standard class value and the classifier-selected class value. The first passage in the figure was manually classified as a relevant passage, and the classifier correctly placed it in the relevant passage class.

```
1  Unidirectional CFRP is discounted  as this splits too easily to be considered a
   realistic alternative   From these findings  the question arises how much the additional
   microstructural feature of a property gradient across the section  which is peculiar to
   palm stems and bamboo culms  contributes to the stiffness and their overall mechanical
   efficiency   This is explored next      1    1    1.000    1.000
2  A large amount of encased knots degrades the value of yew wood  while the high
   extractives content increases its durability (Wagenfuhr  2000).  The crucial difference
   between both species  however  is their raw density: yew has a high raw density with a
   relatively small gradient from earlywood to latewood  while the spruce density is about
   30% lower  and the gradient from earlywood to latewood is high (Keunecke et al 2009).
   Especially due to these last-mentioned differences  both species are very well-suited
   for deriving fundamental structure-property relationships which establish correlations
   between the mechanical behavior of a material and its structural composition  1    1
   1.000    1.000
3  Additionally  viscoelastic effects in the response of those organic constituents to a
   variety of loading conditions appear to be central to dissipation events   Routes to
   address the role of the thin organic constituents in toughening new classes of ceramic
   organic composites should start with the theoretical and computational approaches that
   have been developed by Buehler et al 2008  adapted to constrained sandwich structures
   that include alternating ceramic and organic layers  1    1    1.000    1.000
```

Figure 9. An example of the classifier's output.

The trained classifiers' results for the test set containing the remaining Mechanical Behavior of Biomedical Materials passages can be found in Table 1. The classifier was expected to do well with the first set of testing documents because the documents came from the same set of articles as the training documents so there is overlap between authors and principles in the two document sets. Table 1's columns display the manually identified gold standard classification. The rows represent the classifier's result. To improve the precision of the text mining tool, the irrelevant passages need to be correctly identified by the classifier. If a passage is irrelevant according to the gold standard, but the classifier considers it to be relevant, then the passage cannot be identified and removed from the passage set and the precision will remain low. Although the Naïve Bayes classifier correctly identified all of the relevant passages in the Mechanical Behavior test set, the classifier missed 54% of the irrelevant passages. Since precision is dependent on correctly identifying these irrelevant passages, the maximum entropy (maxent) classifier was then utilized. The maxent classifier was expected to have an improved performance because it is able to take into account feature interaction and it tries to maximize the conditional likelihood of the classes. As seen in Table 1, the maxent classifier did a much better job of correctly classifying the irrelevant passages. Several of the relevant passages were classified incorrectly, which would decrease the text mining algorithm's recall value. While it would be ideal if all of the relevant passages were kept, the small decrease in recall is expected when increasing precision. Since the maxent classifier worked better with the first set of testing documents, the maxent classifier was used for the other testing set as well.

Table 1. Results of the Naïve Bayes (NB) and maximum entropy classifier (Maxent) using the Mechanical Behavior of Biomedical Materials testing documents.

| | | Gold Standard | |
|---|---|---|---|
| | | Relevant | Irrelevant |
| NB Classifier Result | Relevant | 43 | 54 |
| | Irrelevant | 0 | 46 |
| Maxent Classifier Result | Relevant | 41 | 0 |
| | Irrelevant | 2 | 100 |

The trained classifier did not perform as well with the randomized set of testing passages collected from the three journal issues (Table 2). The classifier only correctly categorized 7 of the irrelevant passages, and misclassified 19 of the irrelevant passages. A likely reason that the trained classifier did not perform well with the second testing set is because of the addition of different authors and their writing styles. Since the classifier used the terms from the training documents as features to determine the rules for classification, the trained classifier may have had difficulty with new terms found in the second test set. It is possible that the classifier used the principle terms themselves as markers of the relevant passage class from the training passages. These training document principles may not have appeared in new articles.

Table 2. Results of the maximum entropy classifier on the randomized set of passages from all of the journal issues.

| | | Gold Standard | |
|---|---|---|---|
| | | Relevant | Irrelevant |
| Maxent Classifier Result | Relevant | 23 | 19 |
| | Irrelevant | 0 | 7 |

Both the Naïve Bayes classifier and the maximum entropy classifier had difficulty classifying the testing document passages. The potential for many false negative scores would result in retaining many of the irrelevant passages, keeping the precision score low. The maximum entropy classifier is expected to perform well when a majority of the principles and writing styles are accounted for in the training documents. However, since the algorithm should be able to identify new principles as they appear in the articles, the supervised machine learning classifiers are not useful for this work's text classification purposes.

### Statistical Features Study

The most common feature for text classification is term frequency, as considered in the Machine Learning Classifier Study. Another approach to analysis is the use of document statistics as features. Nonsemantic quantitative features that may be statistically analyzed include the number of words in a document, the average word length in a passage, the maximum word length, the location of a term in the document, and the amount of whitespace [83, 84]. For this study, the simple statistical features used were the number of words per property sentence and the number of words per passage. The research approach and the hypotheses are reviewed, followed by the evaluation results and discussion.

### *Statistical Features Research Approach*

To determine if a difference exists in the number of words per sentence or number of words per passage used between the relevant and irrelevant classes, a subset of the passages from the Mechanical Behavior of Biomedical Materials issue was analyzed. All 58 relevant passages and 86 irrelevant passages for the property of strength were examined. A word count was completed for the property sentence and for the full passage for each of the passages. An unpaired samples t-test was carried out for each feature to compare the irrelevant and relevant passage classes.

The hypothesis is that the relevant passages would have a significantly greater number of words per property sentence. The relevant passages' property sentences need to describe how the property changes and explain the effects. Irrelevant passages that report experimental procedure, list equations and variables, or record the results can be more easily expressed in a concise manner and written with shorter sentences. The passage length hypothesis follows the same line of thought with the expectation that the relevant passages have more words than the irrelevant passages. The relevant passages need to describe the structure-property relationship, which are expected to require additional descriptor words to explain a more complex idea.

*Statistical Features Evaluation and Discussion*

For sentence length, there was not a statistically significant difference in the number of words per property sentence for the relevant passages (M = 30.1, SD = 11.6) as compared to the irrelevant passages (M = 31.7, SD = 11.2), as seen in Figure 10; t(142) = 0.80, p = 0.42. There was also not a statistically significant difference in the number of words per passage; t(142) = 1.59, p = 0.11. On average, the relevant passages had 76.9 words per passage with a standard deviation of 20.6, while the irrelevant passages had a mean of 82.7 words per passage with a standard deviation of 22.4 (Figure 10).

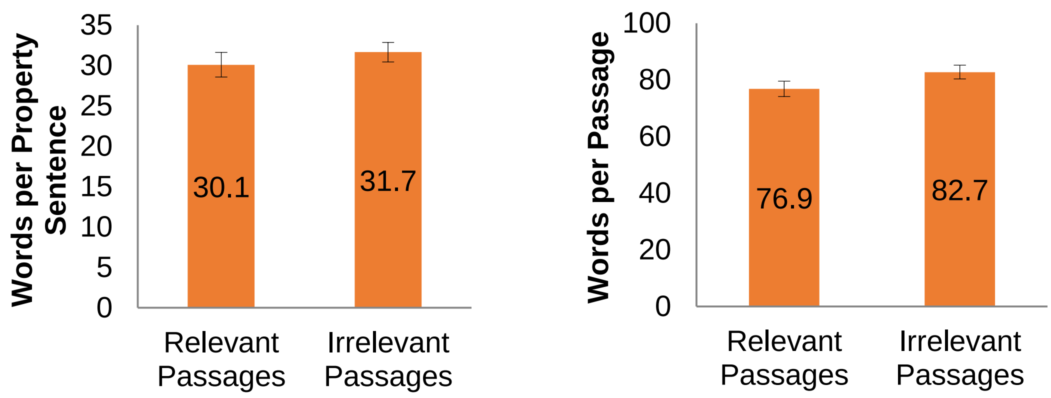



Figure 10. A comparison of the average number of words between the relevant and irrelevant passages by sentence (left) and by passage (right). Error bars show ±1 standard error.

The lack of difference between the relevant and irrelevant passage classes could be due to the writing style of the authors. If an author tends to write in longer sentences, they would likely use many words in the procedural descriptions as well as in the discussion of the results. Author-dependent writing styles can explain the large standard deviations in the results. Another difficulty in distinguishing relevant and irrelevant passage differences is that passages are property-dependent. The same passage may be considered relevant for one property, but irrelevant for another. For example, the passage:

> *"The bone plates have a circulatory blood system that attaches to the animal through veins. This unusual structure is similar to last tiles used on tactical vehicles fig 2b which are hard ceramic hexagonal blocks covered by a polymer. This indicates that the armadillo shell should have excellent impact resistant properties"* [85]

would be considered irrelevant for the property of hardness, but the structure of hexagonal blocks covered by a polymer would be useful for improving toughness. Because the same passage would appear in each class when considering all properties, statistical feature differences are difficult to determine.

## Part-of-Speech Study

Looking towards natural language processing, the next text classification feature considered was part-of-speech. The parts of speech are categories of words that have similar grammatical properties, such as nouns, verbs, adjectives, and adverbs. Part-of-speech can be used in text classification because the different types have different purposes. For the part-of-speech study, the words in a passage are classified into their respective syntactical category, and the overall trends in the relevant and irrelevant passages are drawn.

*Part-of-Speech Research Approach*

The passages used for the statistical features study were again used for the part-of-speech study. The Mechanical Behavior of Biomedical Material issue's passages were run through the Stanford Part-Of-Speech (POS) Tagger [86]. The POS Tagger is a log-linear tagger implemented in Java that automatically tags each word of the passage text with its Penn Treebank English POS tag. The Penn Treebank POS tagset has 45 tags [87], which were more generally categorized into noun, verb, adjective/adverb, and conjunction/preposition groups, following Glier's work [84]. For example, the tag "NN" corresponds to a singular or mass noun, while the tag "NNPS" corresponds to a plural proper noun, but both fall within the general noun grouping. The Noun group consisted of the tags: NN, NNS, NNP, NNPS, PRP, PRP$, WP, WP$, while the Verb group included: VB, VBD, VBG, VBN, VBP, VBZ. The Adjective/Adverb group encompassed the JJ, JJR, JJS, RP, RBR, RBS, and WRB tags, while the Conjunction/Preposition group had the CC and IN tags. To compare the relevant and irrelevant passages for differences, counts for each of these general tag groups were completed.

The hypothesis is that there should be more adjective/adverb tokens and verb tokens in the relevant passages than the irrelevant passages. Adjectives and adverbs are modifiers that are used to add description and detail. Passages containing the structural design principles need to describe the structure in detail so more adjectives and adverbs should be used. For example, in the passage printed in Chapter 1, the sacrificial domain structure is described as consisting of "loosely packed modules". Verbs signify actions and functionality so they should be used to describe how the material property changes and how it is affected by the structure. Once again looking to the previous passage, verbs are used to show that the unraveling of the modules "led to…an increase in the energy absorbed…" and that "similar sacrificial connective proteins have been proposed as toughening agents…" [26].

### *Part-of-Speech Evaluation and Discussion*

The 58 relevant passages had a total of 4458 words, while the 86 irrelevant passages had 7109 words. As seen in Table 3, the most common general tag group were nouns for both the relevant passages and irrelevant passages, followed by conjunctions/propositions, adjectives/adverbs, and then verbs. The "other" group consisted of cardinal numbers, determiners, existential there, predeterminers, possessive endings, to, interjections, and wh-determiners. The percentage of each part-of-speech group was calculated by dividing the word count for that group by the total number of words in the passages in the class. Looking at the percentages, there was not a significant difference between the relevant and irrelevant passages (Figure 11). A two percent difference between the relevant passage class and irrelevant passage class is not enough to accurately classify an 80-word passage with a large standard deviation. Therefore general part-of-speech cannot be used to sort the passages into "relevant" or "irrelevant" classes. Once again, the lack of difference between the two classes may be due to author writing preferences. The difference may also be due to the long multi-sentence passages, which mitigates the effect from the short structure-property relationship description.

Table 3. Counts for the different general part-of-speech tag groups in the relevant and irrelevant passages.

| | Noun | Verb | Adj/Adv | Conj/Prep | Other |
|---|---|---|---|---|---|
| Relevant Passages | 1371 | 580 | 622 | 790 | 1095 |
| Irrelevant Passages | 2317 | 985 | 1059 | 1274 | 1473 |

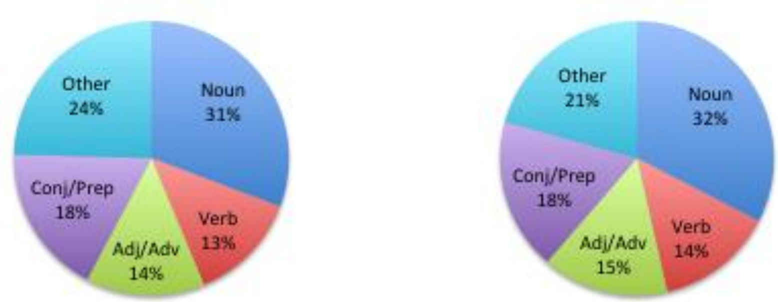



Figure 11. The percentages of the part-of-speech tag groupings in the relevant (left) and irrelevant (right) passages.

## Conclusion

The chapter began with the previously developed Stage 1 algorithm that extracts biological materials' structural design principle passages from materials papers. The text mining tool had a high recall, but a lower precision. The goal for this study was to use classic text classification techniques to categorize the passages into relevant and irrelevant classes in order to remove the irrelevant passages and improve precision. In Stage 2, several approaches were taken for text classification including the use of the machine learning classifiers, statistical features, and part-of-speech analysis. The use of classic text classification techniques with a variety of features such as term frequency, property sentence length, passage length, and parts of speech, were not successful in classifying passages into “relevant” and “irrelevant” classes. Supervised machine learning methods and previously developed algorithms were not able to automatically distinguish differences between the relevant passages and irrelevant passages. It is necessary to concentrate more on the patterns of the words used in the passages to create a rule-based classification scheme rather than utilizing popular tools and statistical features.

# CHAPTER V

# ALGORITHM DEVELOPMENT STAGE 3: MANUAL RULE-BASED CLASSIFICATION

The development of a descriptive bioinspired text mining tool can help bridge the knowledge gap between biological materials research and materials designers to provide the designers with new ideas. The Stage 1 version of the algorithm resulted in high recall, but low precision. Stage 2 explored classic text classification techniques to automatically categorize the passages into a "relevant passage" class and an "irrelevant passage" class. By classifying the passages, the irrelevant passages can easily be removed from the results and improve precision. Since Stage 2's techniques did not prove effective for this dissertation work, Stage 3 of the research involves manually identifying patterns and templates of the relevant and irrelevant passage classes that can be used to sort the passages. Like in Stage 2, an improvement in precision is the chief goal, but recall is still considered in order to return a variety of principles to the tool's users. Manual rule-based classification is not used as often as the Stage 2 approaches for several reasons. First, hand-coded rules are often not able to build upon other's expertise or take advantage of previously developed text classification algorithms. Second, manual identification of the applicable rules and algorithm development is expensive to build and maintain. Manually developed rules also tend to be brittle, meaning they are domain-specific. However, manually developed rules are usually simple and have high accuracy [59, 60].

Manual rule-based classification was used during a challenge that took place in association with the 2002 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD) [88]. The KDD Challenge Cup was a competition where participants were provided a training corpus and testing corpus, and asked to develop text mining systems that determined whether a paper met certain curation criteria and that indicated whether the papers had experimental evidence of gene products. Thirty-two submissions were evaluated. The winning team made use of manually constructed rules and patterns, while honorable mention teams combined manually determined items with automated learning statistics. The contest organizers noted in their evaluation of the competition that solely using automated statistical learning and weighting techniques was not enough for classification, and recommended the manual determination of features to look for and possibly where to find them.

This chapter begins with a discussion of the approach taken to create hand-coded rules. The following sections describe the patterns found in the irrelevant and relevant passage sets. The chapter then reviews the pattern-based revisions made to the algorithm, and an evaluation of the Stage 3 algorithm is completed using a new set of journal articles. The chapter concludes with a discussion of the manual rule-based classification approach and its effects.

## Manual Rule-Based Classification Approach

To identify the patterns, Stage 3 of the research used the passages output from the Stage 1 search tool, as described in Chapter 3. The journal issues used for the passages are the Advanced Materials Journal Volume 21, Issue 4 Biological and Biomimetic special issue [73] and the Materials Science and Engineering C 31 Volume 31, Issue 4, presented at the 6th Symposium on Biological Materials Science at TMS Annual Meeting & Exhibition [80]. Once again, each of the passages were manually sorted into a "relevant passage" class and an "irrelevant passage" class. The relevant passages referred to or described a biological material structure that affects the searched material property. If the passage included a structure that was not from a biological material or described other items that affect the property of interest, the passage was still classified as irrelevant. For example, a passage describing the effect of delamination on the strength and stiffness of an engineered material was classified as an irrelevant passage.

Next, the property search terms and relationship terms in the passages were highlighted to understand how the terms were being used. The irrelevant passages were studied to identify patterns, such as the usage of search terms in unexpected ways. The irrelevant passage patterns were then searched for within the relevant passages. If a pattern appeared often within the relevant passages as well, then the pattern was discarded. Next, patterns were identified for the relevant passages. Hand-coded rules were developed based on the final patterns that were then used to create the Stage 3 algorithm.

An evaluation of the updated algorithm took place using a set of articles that had not yet been tested by the program. The articles were taken from the Mechanical Behavior of Biomedical Material special issue [79]. First, the principles in the articles were manually identified.  Then, full articles were run through the Stage 1 algorithm and the Stage 3 algorithm in order to conduct a comparison of the two.

## Irrelevant Passage Patterns

For the irrelevant passages, analysis was focused on property term usage.  Two main patterns emerged that resulted in a revision to the property term sets and in new exclusion term sets.

The first identified pattern was that some property search terms were consistently not used to describe property effects.  For example, in a materials article, the term “indenter” is almost always associated with the property of hardness.  However, “indenter” is mostly used to describe a hardness test or experimental procedure, rather than how the material’s hardness changes.  Since the term “indenter” was used 18 times in the irrelevant passages, but only one time in the relevant passages, it was removed from the search term list.  Another example term is "strong”.  Rather than referring to a high strength material, the term “strong” is often used in phrases such as “strong association” or “strong ability”, resulting in the removal of the term “strong” from a property search term set as well.  The property search terms were updated with the terms following this pattern.  Since the program has a strict search of the terms, the morphological variations of the search terms were also removed from the search term set.  For example, “indenters” was removed along with “indenter”.  Four property search terms were removed for the property of density, four terms for ductility, twenty-one for hardness, five for stiffness, twenty-four for strength, and six for toughness. The updated property search term sets can be found in Appendix B.

The other identified pattern was the use of property-specific exclusion terms. The exclusion terms are phrases that contain a property search term, but the entire phrase does not describe the property of interest. For example, the phrase "crack density" contains the term "density", which is a desired search term. However, "crack density" does not refer to the density of the material, but rather to the amount of cracks per unit area. Passages describing "crack density" should not be returned when searching for density-related design principles. Another example is "pull off strength". While "strength" should be retained as a search term, the phrase "pull off strength" actually refers to the property of adhesion. All of the passages were inspected to determine if the search terms were used in a common phrase, and if the phrase described the searched property. However, since exclusion terms are property-specific and do not appear consistently in all articles, other text was used to identify additional exclusion terms. The Nature Materials journal, which regularly has one of the highest journal scores in Materials Science, was utilized to identify additional exclusion terms. A total of 46 journal issues and 802 articles ranging from the 2002 Volume 1, Issue 1 through 2006 Volume 5, Issue 6 were examined. A search was completed for each of the property search terms throughout the articles. The exclusion terms were selected based on domain knowledge or on the researched phrase definition. Each property has its own set of phrases to exclude from the property search. Exclusion term phrases that were longer than three words were generally rejected since the longer phrases are often too specific and less commonly used as descriptors. The morphological variations of the phrases were added to the exclusion term list so "crack densities" was included as well. Using the passages and the Natural Materials journal issues, density had 267 exclusion terms, ductility had 3, hardness had 95, stiffness had 68, strength had 142, and toughness had 48 (Appendix B). The property-specific exclusion terms were added to the program as a filter to screen passages that do not describe the user-selected properties.

**Updated Algorithm and Preliminary Evaluation with Irrelevant Passage Patterns**

The patterns identified in the irrelevant passages were implemented into the updated algorithm (Figure 12). As in the Stage 1 algorithm, the Stage 3 algorithm begins with a biological search. The property search terms were revised to reflect the search terms that were not consistently used to describe property changes. A search for the property-specific exclusion terms was added after the property term search. If the identified property term was part of an exclusion term, then that property term would no longer be considered as a basis for the passage. These patterns were coded into the text mining tool. The identification of a relationship term and the sentence distance requirement between the relationship and property terms remained unchanged.

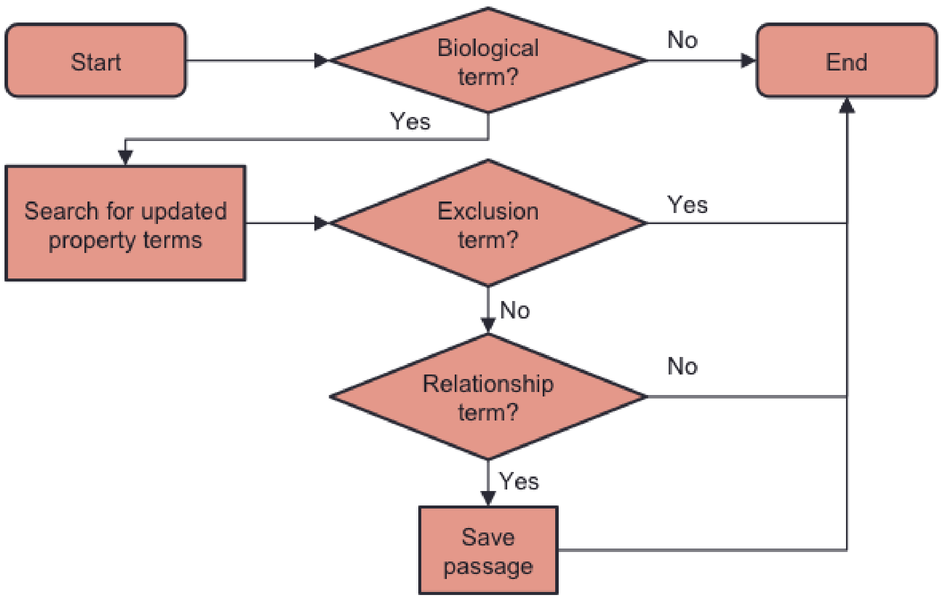



Figure 12. The Stage 3 algorithm updated with the irrelevant passage patterns.

A preliminary evaluation was completed using the algorithm updated with the irrelevant passage patterns and the Advanced Materials special issue and the Materials Science and Engineering issue. The evaluation is considered preliminary because these articles contained the training passages used for the identification of the patterns, which should result in increased precision values. Once again, precision was calculated by dividing the number of relevant passages by the number of passages extracted by the program. As seen in Table 4, precision was improved for each of the individual property searches. The change in the number of retrieved passages due to the updated property search terms was not as significant as the improvement in the number of relevant passages. The overall algorithm precision for all of the properties increased from the

Stage 1's 30% to 53%, which confirmed that the addition of these irrelevant passage patterns aided in the removal of irrelevant passages.

Table 4. The preliminary results after the addition of the irrelevant passage patterns to the Stage 3 algorithm as compared to the Stage 1 algorithm using the training passage articles.

| Properties | Stage 1 Relevant Passages | Stage 1 Retrieved Passages | Stage 1 Precision | Stage 3 Relevant Passages | Stage 3 Retrieved Passages | Stage 3 Precision |
|---|---|---|---|---|---|---|
| Density | 4 | 41 | 10% | 10 | 40 | 25% |
| Ductility | 11 | 25 | 44% | 16 | 25 | 64% |
| Hardness | 8 | 38 | 21% | 17 | 36 | 47% |
| Stiffness | 29 | 64 | 45% | 54 | 86 | 63% |
| Strength | 7 | 51 | 14% | 27 | 53 | 51% |
| Toughness | 16 | 28 | 57% | 34 | 59 | 58% |
| **Total** | **75** | **247** | **30%** | **158** | **299** | **53%** |

### Relevant Passage Patterns

Two relevant patterns were identified through the examination of the relevant passages. The first observed pattern was that a descriptor of how the property changes was often found in the same sentence as the property search term. The term describing the property change is referred to as the "identifier term". Thus the first relevant passage pattern was the usage of the property term and identifier term in the same sentence. An example of this pattern can be found in the sentence:

> *"The orientation of the channels that run parallel to the long axis of the teeth probably increase their bending stiffness…" [89]*

The passage's channel orientation principle likely causes an increase (identifier term) in the material's stiffness (property term). The advantage of using this pattern is that it helps to exclude experimental descriptions such as:

> *"the known value of the elastic modulus… as determined by nanoindentation with a commercial instrument triboindenter hysitron inc…" [85]*

Two types of identifier terms were found. The first type shows a comparison or change of the property such as "lower", "more", "reduces", or "differ". The second type of identifier term shows the cause of the property change such as "due to", "function of", or "depended on". All of the relevant passages were examined to find the identifier terms, and a total of 88 terms were found. Due to the strict nature of the search, the morphological variations of the terms were also added. For example, along with the term "increase", the terms "increased" and "increasing" were included in the identifier term list. This resulted in a total of 172 identifier terms, which are displayed in Appendix C. The identifier and property pattern appeared in 82% of the relevant passages.

Many of the relevant passages that did not follow the identifier and property term pattern instead used a "property-specific stand-alone term". These stand-alone terms are property terms that do not require an identifier term. A property-specific stand-alone term is used in the passage:

> *"hence by maximizing imin, a triangular column makes use of material in a better fashion to yield stiffer structures" [75]*

The property term "stiffer" is a comparative term that does not require a separate term to quantify it. The stand-alone terms make use of several different suffixes. They include the property terms ending with "-er" and "-est" for comparison. The terms also have the suffixes "-en", "-ened", and "-ening" such as "stiffen", "stiffened", and "stiffening". These stand-alone terms can be found in Appendix C. Property-specific stand-alone terms appeared in 14% of the relevant passages.

The final pattern identified in the relevant passages was the unnecessary length of the passages based on the property term and structural design principle locations. The original algorithm allowed for the property and relationship terms to appear within two sentences, and then added two sentences before and after the search terms to form a passage. It also combined passages that were next to each other with the assumption that the passages would be discussing the same topic. However, this assumption resulted in passages that were up to a page and a half in length, while only one or two sentences were actually of interest. The long passages may also have retrieved principles coincidentally due to the extraneous sentences, which would artificially inflate the original recall and precision values. Shorter passages are advantageous because they have less unrelated text so the tool's user will need to spend less time reading through the passages. The shorter passages would also be useful for passage grouping because the key principle terms would represent a greater proportion of the total word count. To determine the new passage length, all of the relevant passages were examined for their property term and principle structure term locations. Table 5 displays the location of the principle terms as compared to the property search terms. The property term and structural design principle appeared in the same sentence 79% of the time in the relevant passages. The new passage length was chosen to be three sentences to include the sentence before and after the property term, capturing 95% of the relevant passages.

Table 5. The distance of the design principle terms from the property search terms.

| | 2 sentences before | 1 sentence before | Same sentence | 1 sentence after | 2 sentences after | Total Relevant Passages |
|---|---|---|---|---|---|---|
| Density | 0 | 1 | 9 | 1 | 0 | 11 |
| Ductility | 0 | 2 | 19 | 1 | 0 | 22 |
| Hardness | 2 | 2 | 15 | 0 | 1 | 20 |
| Stiffness | 2 | 8 | 80 | 14 | 3 | 107 |
| Strength | 0 | 1 | 30 | 2 | 1 | 34 |
| Toughness | 2 | 5 | 43 | 3 | 1 | 54 |
| Total | 6 | 19 | 196 | 21 | 6 | 248 |
| Percentage | 2% | 8% | 79% | 8% | 2% | |

**Updated Algorithm and Preliminary Evaluation with All Passage Patterns**

The final updated algorithm, which takes into account both the relevant passage and irrelevant passage patterns, can be seen in Figure 13. Similar to the irrelevant passage pattern algorithm, the program uses the updated property search terms, and then checks that the property term is not part of an exclusion term. If the property term is not part of an exclusion term, a check is made to determine if the property term is a stand-alone term. If it is, the passage is saved with the updated three-sentence length. If the property term is not part of a stand-alone term, then the program searches for an identifier term within the same property term sentence. If an identifier term is found, the passage is saved with the new passage length.

One key difference between the Stage 1 algorithm and the final Stage 3 algorithm is that the original relationship term search has been replaced with the identifier term search. When analyzing the irrelevant passages, the relationship terms were found to appear often, up to 20 times per passage. The terms were too vague, resulting in the relationship terms being used more often and for different purposes than expected. The passages containing the stand-alone terms were examined with the updated patterns, and all of the passages were still extracted without the use of a relationship term. Of the 170 relationship terms, 58 overlapped with the identifier terms. The overlap came from the identifier terms that describe the causes of property change, and not from the identifier terms defining the quantitative change. Rather than completing a relationship search followed by an identifier search, the relationship search was replaced. The advantages of replacing the relationship search are that the algorithm is faster with only one round of searching and is easier to implement.

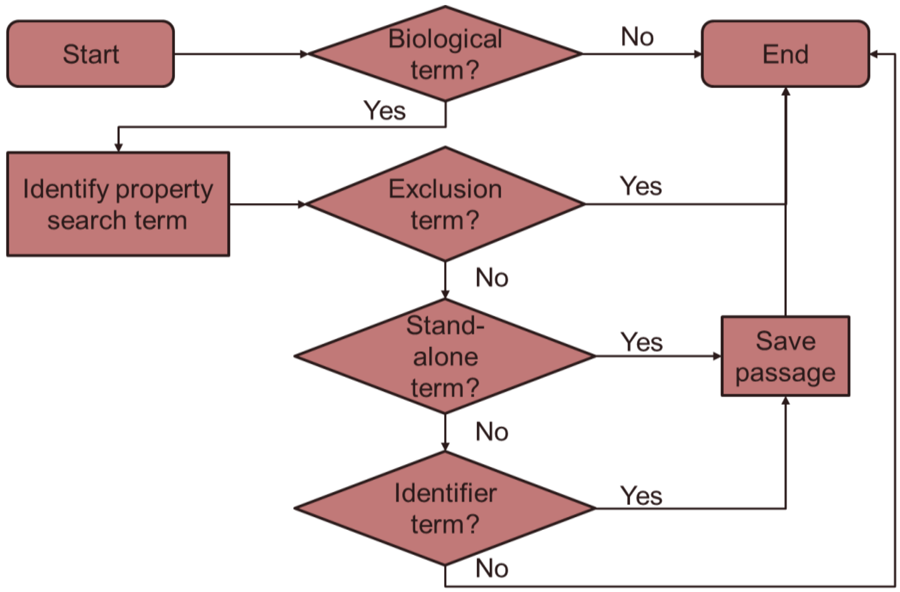



Figure 13. The Stage 3 algorithm updated with both the relevant and irrelevant passage patterns.

Once again, a preliminary evaluation of the algorithm was completed using the training passage articles used to identify the patterns, which likely inflated the precision results. Comparing Table 6 to Table 4, the tool's precision was further increased with the addition of the relevant passage patterns. However, the percentage of principles identified (recall) decreased from 91.6% in Stage 1 to 77.4% in Stage 3. The lower recall value is still good, especially considering the significant increase in precision. A change in the type of passages returned was observed. As anticipated, there were fewer procedural and general description passages.

Table 6. The preliminary results comparing the original algorithm (Stage 1) to the updated algorithm containing both the irrelevant and relevant passage patterns (Stage 3) using the training passage articles.

| Properties | Stage 1 Relevant Passages | Stage 1 Retrieved Passages | Stage 1 Precision | Stage 3 Relevant Passages | Stage 3 Retrieved Passages | Stage 3 Precision |
|---|---|---|---|---|---|---|
| Density | 4 | 41 | 10% | 11 | 22 | 50% |
| Ductility | 11 | 25 | 44% | 21 | 26 | 81% |
| Hardness | 8 | 38 | 21% | 8 | 16 | 50% |
| Stiffness | 29 | 64 | 45% | 75 | 107 | 70% |
| Strength | 7 | 51 | 14% | 33 | 49 | 67% |
| Toughness | 16 | 28 | 57% | 25 | 39 | 64% |
| **Total** | **75** | **247** | **30%** | **173** | **259** | **67%** |

## Algorithm Evaluation

A full evaluation of the final updated Stage 3 algorithm (Figure 13) was completed using the Journal of the Mechanical Behavior of Biomedical Materials Volume 4, Issue 5 [79]. To more fairly compare the Stage 3 algorithm to the Stage 1 algorithm, the Stage 1 algorithm was modified slightly to also output a three-sentence passage. To accomplish the passage length reduction, the property terms and relationship terms were required to be within the same sentence. The revised Stage 1 algorithm Java code is included in Appendix D. The results of the revised Stage 1 algorithm and Stage 3 algorithm evaluation can be found in Table 7.

One notable result is that the modified Stage 1 algorithm had a higher precision and lower recall than the original Stage 1 algorithm. The original Stage 1 algorithm's precision was approximately 32% and the recall was over 90% for the two journal issues used in the original evaluation. For the Mechanical Behavior of Biomedical Materials journal issue, the original Stage 1 algorithm achieved a recall of 94% for the six individual properties. The revised Stage 1 algorithm's precision increased to 46%, although the recall dropped to 82%. Based on the differences between the two Stage 1 algorithms with the same article set, the change in precision and recall are due to the search term requirements. The increase in precision was surprising considering that the original Stage 1 algorithm's value was likely inflated due to the overly long passages.

The revised Stage 1 algorithm had slightly better performance in both the precision and recall metrics. The Stage 3 algorithm retrieved 80% of the manually identified principles with 45% of the returned passages considered relevant. To put the results in perspective, approximately half of the output passages are potentially useful. From the tool user's point of view, they are able to select the property they want to design for, and essentially every other returned passage may offer a source of inspiration. The results are satisfying, especially considering that the program is searching for domain-specific information that requires expertise.

While the Stage 3 algorithm slightly underperformed for the two metrics as compared to the revised Stage 1 algorithm, it did return fewer irrelevant passages, as seen in Table 7. The Stage 3 algorithm also had a 22% decrease in total number of passages. If the goal is to reduce the amount of reading required by the user to identify the principles, the Stage 3 algorithm would be considered superior.

A final note is that there were differences in the passages returned between the two algorithms. The difference in passages was based on the updated property search terms, along with the addition of exclusion terms and identifier terms. The Stage 1 algorithm identified some principles that were not found by the Stage 3 algorithm, while the Stage 3 identified some new principles as well.

Table 7. An evaluation and comparison of the revised Stage 1 and final Stage 3 algorithms.

| | Stage 1 | | | | | Stage 3 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Properties | Recall (Calc) | Recall (%) | Relevant Passages | Irrelevant Passages | Precision | Recall (Calc) | Recall (%) | Relevant Passages | Irrelevant Passages | Precision |
| Density | 15/19 | 79% | 29 | 42 | 41% | 14/19 | 74% | 26 | 12 | 68% |
| Ductility | 7/9 | 78% | 7 | 50 | 12% | 8/9 | 89% | 5 | 30 | 14% |
| Hardness | 13/19 | 68% | 98 | 92 | 52% | 12/19 | 63% | 32 | 29 | 52% |
| Stiffness | 36/41 | 88% | 128 | 139 | 48% | 38/41 | 93% | 109 | 118 | 48% |
| Strength | 33/40 | 83% | 58 | 85 | 41% | 30/40 | 75% | 65 | 132 | 33% |
| Toughness | 36/43 | 84% | 81 | 71 | 53% | 34/43 | 79% | 70 | 57 | 55% |
| Total | **140/171** | **82%** | **401** | **479** | **46%** | **136/171** | **80%** | **307** | **378** | **45%** |

## Conclusion

Stage 3 presented an approach to manually identifying patterns of the irrelevant and relevant passages in order to produce a rule-based classification system. The irrelevant passage patterns resulted in the removal of selected property search terms, as well as the addition of property-specific exclusion terms to the program algorithm. The relevant passage patterns consisted of the usage of an identifier term and a property term in the same sentence and the usage of a property-specific stand-alone term. The patterns identified from the passages can be extrapolated to apply to additional articles and authors. The patterns were used to create the Stage 3 algorithm, and the algorithm was evaluated against a revised Stage 1 algorithm. The Stage 1 algorithm had slight revisions for a more fair comparison of the passages between the two algorithms.

The changes made to the original Stage 1 algorithm resulted in an increase in precision, as desired, which also corresponded to a decrease in recall. By simply updating the passage length and the requirement that the relationship and property term appear in the same sentence, the precision of the algorithm was increased from having one relevant passage out of every three passages to nearly one relevant out of every two. The Stage 3 algorithm had slightly lower recall and precision scores, but its decrease in the number of passages allows the program user to look through the results more quickly. With fewer irrelevant passages, the user can spend less time reading unsuitable passages.

Based on the program user's goals, either stage may be chosen. If the user wants more sources of inspiration and more principles, Stage 1 may be desired. Whereas if the user just wants to decrease the amount of time required to identify some principles, Stage 3 would be more useful. Either way, for practical application, biological materials' structural design principles are being identified within the output passages.

# CHAPTER VI

# ALGORITHM DEVELOPMENT STAGE 4: PRECISION CLASSIFICATION

Extracting biological material structural design principles from a materials corpus has been found to be difficult because the structures and principles are unknown prior to searching. In Stage 1, an algorithm was created that was able to extract passages with structural design principles. However, the next issue that arose was the many irrelevant retrieved passages due to mentions of material properties without reference to material structure. To alleviate the issue, text classification was examined to classify the passages into "relevant" and "irrelevant" classes. Stage 2 noted that the unknown material structures, along with the variability in the English language and author writing styles, make the utilization of classic text classification techniques difficult. In Stage 3, an updated version of the structural design principle passage-retrieval algorithm was created using manual rule-based classification, which resulted in improved precision and a reduced number of passages while retaining high recall. Stage 4's final algorithm goal is to improve precision in order to further reduce the effort needed by a materials designer to identify structural design principles of biological materials. The final iteration of the algorithm aims to more strictly filter out the irrelevant passages that do not describe the structure-property relationships. It is recognized that the improvement in precision will be accompanied by a decrease in recall. Between the Stage 4 algorithm and the Stage 1 revised algorithm, both sides of the precision and recall trade-off will be explored. A user can then select which version of the tool to use based on which metric they want to emphasize. The Stage 4 algorithm is to be used when the materials designer wants to quickly search for structural design principle options, but may not require an exhaustive investigation.

The initial Stage 1 version of the tool identified over 90% of the principles (recall), although only approximately 32% of the returned passages were relevant (precision) [90]. Stage 3's manual rule-based classification update resulted in updated search terms, and the addition of property-specific exclusion terms, stand-alone terms, and identifier terms in a three-sentence passage. In comparison, the Stage 3's algorithm resulted in a recall of 80%, but improved precision of 45%. The Stage 1 algorithm was also revised to output three-sentence passages for easier program comparison to Stage 3's algorithm. The revised Stage 1 algorithm achieved a precision of 46% and a recall of 82%, which was similar to other initial biomedical text mining classification algorithms [67].

This chapter covers several different approaches taken for the final stage of algorithm improvement. First, another round of manual rule-based classification is completed to identify stricter patterns and templates for the Stage 4 algorithm. Next, a more in-depth part-of-speech study is conducted on the identified pattern terms in an effort to use natural language processing to classify the text beyond the search and template-based methods. After, the results of a passage activity are presented. Four graduate-level mechanical engineering students analyzed passages individually in order to elucidate new rules for the updated algorithm. The tool is updated to take into account the additional patterns and rules, and a final program evaluation is performed using a new set of articles.  This chapter concludes with a comparison between the revised Stage 1 algorithm and the final Stage 4 algorithm.

## Manual Rule-Based Classification Round II

Similar to the Stage 3 research, manual rule-based classification was used to identify patterns and templates of the relevant and irrelevant passages.  The discovered irrelevant passage rules can then be applied when searching for passages in order to filter them from the results.  Even though manual rule-based classification algorithms are expensive to build and maintain, Stage 4 began with manual rule-based classification the technique because the method can achieve good scalability and high accuracy if the rules are created by a domain expert [59, 60].  Previously, the identified relevant and irrelevant passage patterns were more general in order to preserve high recall and retain the relevant useful passages.  For Stage 4, the rules are to be stricter in order to improve precision, with the recognition that some relevant passages may be lost.  This section begins with an overview of the Stage 4 manual rule-based classification approach.  Next, a closer examination of the irrelevant passage class's research approach and the resulting irrelevant passage patterns are presented.  The approach to identify patterns and templates within the relevant passage class is then described, followed by the resulting relevant passage patterns.  These patterns will be applied to the updated Stage 4 algorithm.

*Manual Rule-Based Classification Research Approach*

For manual rule-based classification, the passages returned from the Stage 3 algorithm were used to identify the Stage 4 patterns and templates. The passages were from the Journal of the Mechanical Behavior of Biomedical Materials Volume 4, Issue 5 Special Issue on Natural Materials / Papers from the Third International Conference on the Mechanics of Biomaterials and Tissues [79]. Manual passage sorting was completed to classify the passages as relevant or irrelevant. The relevant passages were required to reference a structural design principle for the searched property.

For both the relevant passage class and the irrelevant passage class, all of the search terms, which consisted of the property, identifier, exclusion, and stand-alone terms, found in the Mechanical Behavior of Biomedical Materials journal issue passages were highlighted. For the property and stand-alone property terms, the number of times each of the terms were used in the relevant passage class and the irrelevant passage class were compared. Although the property terms were determined to be sufficient to identify a majority of the structural design principles in Stage 1, an evaluation was completed to ensure that all of the property terms were useful. A check was made that the appearance of the terms in the irrelevant passage class did not significantly outnumber those in the relevant passage class. The largest discrepancy in term usage were for the terms “modulus” and “density”, which each appeared 26 times more often in the irrelevant passages. “Modulus” appeared 140 times in the irrelevant passages and 114 times in the relevant passages, while “density” appeared irrelevantly 32 times and relevantly 6. However, since both terms were used to identify relevant passages and each are part of numerous property-specific exclusion terms, the terms were kept. After consideration, all of the remaining property search terms were retained as well. A similar process was completed for the identifier terms. Additional exclusion terms were identified from the irrelevant passages, and added to the exclusion term sets.

Next, patterns were identified within the two passage classes. The processes specific for the irrelevant and relevant passages will be further described in the Irrelevant Passage Research Approach section and the Relevant Passage Research Approach section, respectively. After identifying the new patterns, a comparison was completed between the two classes. The usage of the relevant passage patterns were identified within the irrelevant passages, and the irrelevant passage patterns were found in the relevant passages. The results were used to determine if the patterns should be added to the Stage 4 algorithm, and if so, if the pattern should be applied by property sentence or by passage. Finally, an investigation took place to ensure that there was no overlap between the terms in the different pattern term sets.

### *Irrelevant Passage Research Approach*

Three rounds of passage investigation were needed to select the rules identifying irrelevant passages. The first two examinations used the irrelevant passages from the Journal of the Mechanical Behavior of Biomedical Materials journal issue from Stage 3. While reading through the irrelevant passages, potential patterns were first identified based on the usage of non-alphabetic characters and passage topic. In the second round, the irrelevant passages were studied more closely to create lists of potential terms based on the irrelevant topics, and exceptions to the patterns were noted.

In the final round of investigation, the usage of the patterns in the irrelevant and relevant passage classes were compared in order to select the final irrelevant passage patterns, the patterns' required terms, and if the pattern requirements should be by sentence or by passage. To aid in pattern selection, an additional set of relevant passages were manually identified within the Journal of the Mechanical Behavior of Biomedical Materials journal issue. The manually selected passages were required to specifically describe the structure-property relationship. Although the passages described relationships for all six of the program-included properties, the passages were not separated by property. All of the passages describing the relationships may not have been identified; however, the 163 manually found passages were just used to support pattern identification.

For the non-alphabetic patterns, the relevant and irrelevant passage comparison was completed using the Mechanical Behavior of Biomedical Materials journal issue's manually identified relevant passages (163 passages) and the Stage 3 irrelevant passages for strength (71 passages) and stiffness (118 passages). If the pattern appeared more often in the irrelevant passages, then the pattern location was considered. The percentage of false negatives and true negatives resulting from the patterns by the middle property sentence and by the full passage were compared. A pattern requirement decision was then made based on if the removal of irrelevant passages would outweigh the number of relevant passages that would be missed. A detailed example of this decision can be found in the Irrelevant Passage Patterns section with the final selection of irrelevant passage patterns.

For the patterns that required term sets, an additional decision was needed to determine the set of terms. Beyond the manually identified relevant passages from the Mechanical Behavior of Biomedical Materials journal issue, the stiffness and toughness relevant passages were taken from the Advanced Materials Biological and Biomimetic special issue [73] used for testing in Stage 1. The irrelevant passages consisted of the stiffness and toughness passages from the Mechanical Behavior of Biomedical Materials journal issue and the Advanced Materials Biological and Biomimetic special issue. A total of 217 relevant passages and 215 irrelevant passages were utilized. A similar selection process to the non-alphabetic patterns was conducted, followed by the pattern term selections. The number of times a potential term was used was normalized by the number of passages in the relevant or irrelevant class. Based on the differences in the classes, a cut-off point was selected to determine which terms to retain for the final pattern. Once again, an example of this decision is in the Irrelevant Passage Patterns section.

### *Irrelevant Passage Patterns*

**S**ix final irrelevant passage patterns were identified. A discussion of the pattern requirement selections, exceptions to the patterns, and term determination are included in this section.

The first irrelevant passage pattern is the usage of numerical figures. Based on a study of the irrelevant passages, numbers are often used to describe procedures or list results, rather than discussing the cause of property changes. An example of this pattern can be found in the following sentence:

> *"In Type-I collagen the weight of each of the three helices is 140 kDa with the length of 300 nm that gives a line density of* $\lambda = (3\times 140000)/300 = 1400$ *Da/nm."* [91]

There are a variety of noted exceptions to the numerical figure pattern. Years (1991 or 2002ab), chemical formulas ($H_2SO_4$), and paper section numbers (1. Introduction) should not result in the passage being removed from the output. Passages containing figure, table, or equation numbers should also be allowed. Percentages are not included in the pattern since they are often used to describe how much a property changes, which is desired in a passage. Finally, numbers written out in words are not part of the pattern as they may be used to quantify the number of items or structures. By middle property sentence, the pattern was found to appear 22.2% of the time in irrelevant sentences (true negatives) and 8.6% of the time in relevant sentences (false negatives). By passage, the pattern correctly identified 34.4% of the true negatives, which misidentifying 17.8% of the false negatives. Based on the sharp increase in false negatives, which would lead to a significant loss in relevant passages, the numerical figure pattern was chosen to be used by sentence. If a numerical figure appears in the sentence prior or after the property sentence, the passage can still be considered relevant.

The next pattern is the use of ratios and proportions within the irrelevant passages. Most of the time, the forward slash is used to define ratios for equations (t1/l1), describe testing rates (mm/s), or as part of the results' units ($MJ/m^3$), which are not useful in identifying the structure-property relationship. However, forward slashes are allowable when they are used between words such as "and/or", "theory/calculations", and "bending stiffness/mass". By property sentence, the forward slash irrelevant passage pattern was found to occur in 1.1% of the irrelevant sentences and 1.2% of the relevant sentences. Since these results were similar, the pattern was set to correspond to full passages. By passage, there were 5.3% true negatives and 2.5% false negatives. An example of a true negative passage with the forward slash pattern can be seen as follows:

> *"The ratio of the mechanical properties of the tooth enameloid relative to the ganoine layer of the scale model is henceforth defined as the mechanical ratio, <u>ME/G</u>, and is applied to both the modulus and the yield stress (i.e., <u>ME/G</u> = modulus ratio = <u>EE/EG</u> = yield stress ratio = <u>σY,E/σY,G</u>)." [92]*

The third pattern for irrelevant passages is the usage of the word "not" before the relationship term, indicating a relationship does not exist. The pattern appeared four times more often in the irrelevant passages than the relevant passages. An example of this pattern can be found in the following sentence where "depend" is the relationship term:

> *"…just as in metal physics Young's modulus does <u>not depend</u> much on the grain structure." [91]*

The fourth pattern is based on the passage topic. The passages describing the method or procedure used should be removed since sample preparation and testing do not discuss the property effects. Indicators of method passages include terms such as "extensometer", "frequency", "load application", "measurements", and "rate". The method pattern can be seen in the following sentence:

*"The displacement <u>rate</u> range used was between 10−4 mm/s and 10 mm/s corresponding to strain <u>rates</u> 10−5 s−1 and 1 s−1, respectively." [93]*

The method pattern requires a set of corresponding method terms. During the second round of analysis using the irrelevant passages from the Mechanical Behavior special issue, 50 potential terms were identified. Then, the three relevant passage sets and four irrelevant passage sets, mentioned in Irrelevant Passage Research Approach section, were used to count the number of times each term appeared. After normalizing the counts by the number of passages, the method terms appeared in an average of 2.5% of the irrelevant passages and 1.4% of the relevant passages. Based on these percentages, the terms that appeared at least 1% more often in the irrelevant passages were kept, although generally the difference in the selected terms was greater. The final set of 27 method terms can be found in Appendix E. Using the final set of terms, the property and passage comparison was completed. When the method pattern was used by passage, there was a very high identification of 89% of the true negatives, and 28% identification of false negatives. The pattern usage by passage had a significant improvement in classification as compared to the usage by sentence, which only correctly identified 43% of the irrelevant sentences, while incorrectly identifying 9% of the relevant sentences. Due to the significant increase in true negative detection by passage, passages that contained a method term are to be excluded from the final returned set of passages.

The fifth pattern is the usage of calculation terms. These include phrases to define equation variables such as "hardness h", "modulus e", and "stress y"; as well as terms used to explain how calculations are completed such as "calculated", "compute", and “determined by”. The same process used for method term selection was used to determine the calculation terms. The final 25 calculation terms can also be found in Appendix E. The following is an example of the calculation pattern:

> *“The hardness H of the layers, subtracting the effects of the Ti substrate, can be calculated by the Bhattacharya and Nix analytical model (1988), given by … where E is the elastic modulus, Y the yield strength, h the contact depth, and t the layer thickness.” [94]*

For the phrases defining variables in equations pattern, passage examination included all of the Mechanical Behavior journal issue irrelevant passages because these phrases are property-specific. The other calculation terms were selected using the same procedure as the method pattern. The middle sentence usage of the pattern resulted in correctly identifying 6.51% of the irrelevant sentences and incorrectly identifying 0.92% of the relevant sentences. By searching for the calculation terms within the entire passage, 21.86% of the irrelevant passages could be removed, while also removing 6.45% of the relevant passages. Since the Stage 4 algorithm was focused on improving precision, the utilization of the pattern by passage was chosen for the updated algorithm.

The final irrelevant passage pattern is the usage of result terms. Result passages either list the results or the analysis needed to determine the results, rather than describing how properties change or their relationship to a structure. The result terms include “assumed”, “curves”, “slopes”, and “values”, with a total of nine final terms (Appendix E). The following passage contains the result pattern:

> *“At the lowest strain rate (10−5 s−1), the value of the elastic modulus was 12 GPa, whereas at the strain rate of 10−3 s−1 and above the elastic modulus was 21 GPa.” [93]*

By sentence, the result terms appeared in 50% of the true negatives and 15% of the false negatives.  By passage, the terms appeared in 179% of the true negatives and 52% of the false negatives.  The true negatives passage value was greater than 100% because the counts were normalized by the number of passages, and the individual passages contained multiple result terms.   Due to the high appearance of result terms in the false positive passages, the result pattern was selected to be used on the property sentence level.

Multiple patterns can appear within one passage, as can be seen in the given example passages.  The patterns were developed to identify certain categories of irrelevant passages, such as the methods and the results.  Passages within these categories have multiple similarities, which translate to multiple patterns.

### *Relevant Passage Research Approach*

For the relevant passages, templates for the passages were explored as opposed to the usage of specific characters or terms.  Four rounds of passage investigation took place to identify the relevant passage patterns.  Round one consisted of looking through Stage 3's Mechanical Behavior special issue relevant passages to first identify potential patterns.  The manually identified relevant passages from the journal issue were also examined to ascertain why certain passages were not identified, leading to the creation of new patterns.  The second round consisted of examining the passages more thoroughly and noting down possible terms for the pattern list sets.  Relevant passages from the Advanced Materials and Materials Science and Engineering special issues were also used to identify terms. In the third round, pattern term selection took place. The number of appearances of the terms were counted in the relevant passages and the irrelevant passages. It was discovered that the terms that appeared more often in the irrelevant passages than the relevant passages were still used often within the relevant passages. For example, the term that had the largest discrepancy between irrelevant and relevant uses was the term "most". The term appeared in 61 of the irrelevant passages, but also appeared in 43 of the relevant passages. Since the term was still used often in the relevant passages, it was saved as a relevant passage term. The terms that were still used often in the relevant passages were saved with the belief that they would improve recall. The irrelevant passages would then be removed using filters built from the patterns described in the previous section. The fourth examination compared the pattern template usage in the relevant and irrelevant passages from the Mechanical Behavior special issue's revised Stage 1 results. The six individual properties were examined for the relevant passage templates, which will be described in detail in the Relevant Passage Patterns section. By implementing the pattern templates, 64.1% of the relevant passages were retained, while avoiding 61.1% of the irrelevant passages. All of the pattern templates will be used in the final algorithm because they all appeared in passages within the relevant passage class, but none consistently appeared together.

### *Relevant Passage Patterns*

Stage 4's relevant passage patterns were in the form of passage templates. There are six templates that make use of four different term sets. The term sets are the modification, causation, correlation, and comparison terms. The modification term set describes how the property is modified and changed, and includes terms like "increase", "decrease", and "reduce". The modification terms are similar to the Stage 3 identifier terms in that they describe a change in property, but they are newly selected and more clearly defined. The causation terms are used to indicate that a causal relation exists within the passage by using phrases like "caused", "due to", "result of", and "if … then" statements. These causation terms are similar to the Stage 1 relationship terms and the Stage 3 identifier terms that describe the cause of a property change. The third set of terms are the correlation terms, which demonstrate a correlational relationship. The correlation terms include "correlated to", "dependent on", "link between", and "ruled by". The fourth, and final, set of relevant passage terms are the comparison terms. Comparison terms indicate differences in results based on features using phrases such as "compared to", "in comparison", and "in contrast with". The 61 modification terms, 100 causation terms, 16 correlation terms, and 11 comparison terms are included in Appendix F.

Six pattern templates were selected for the relevant passages. The first template is the usage of a causation term, modification term, and property term within a passage. The passage should thus indicate that a causal relation exists and how the property is changed. Since the causation term is expected to appear with the material structure information, the causation term may appear anywhere in the passage. The modification term describes the property change so it must appear within the property sentence. An example passage based on the first template follows:

> *"Therefore, for birds capable of flight, temporal effects and fiber alignment gradients from the proximal to distal end may contribute to an increase of at least 100% in stiffness or a decrease in failure strength by more than 200%." [95]*

The second pattern is similar, and makes use of a causation term and a stand-alone property term. The stand-alone terms identified in Stage 3 take into account the property effects so the stand-alone term is able to replace both the property term and the modification term. The property-specific stand-alone terms include words such as "stiffening" or "stiffest".

The third template describe passages that include a correlation term and a property term. Passages containing the correlation term directly state that a correlational relationship exists between the structure and property, and may not describe how the property changes. Since the correlation term is replacing the causation and modification term, it must appear in the same sentence as the property term. The pattern can be seen in the following passage:

> *"More recently, Sevostianov and Kachanov (2000) studied a link between structure of a porous space of the osteonal cortical bone and the bone's anisotropic elastic moduli." [93]*

In passages with the fourth pattern, a correlation term appears in the same sentence as a stand-alone property term. Like the third pattern, a modification term is not required within the passage so the stand-alone property terms act as a standard property term.

The usage of a comparison term, modification term, and property term compose the fifth pattern template. Rather than directly stating the causal or correlational relationship between the structure and property, the comparison term is used to compare property effects of two different structures. An example of the fifth pattern is shown in the example:

> *"In general, bone samples with lamellar microstructure maintained higher elastic modulus and ultimate strength, compared to those containing secondary osteons." [93]*

While the example passage is just one sentence, the template is generally used for entire passages. The comparison term needs to appear within the passage, while the modification term must appear in the property term sentence.

The final template is the usage of a comparison term and a stand-alone property term. Once again, the usage of a stand-alone term negates the modification term requirement.

### *Conclusion*

In this section, the second round of manual rule-base classification was described. The approaches to identifying rules that can classify the Stage 3 passages into relevant and irrelevant passage classes were covered. Manual rule-based classification resulted in six new irrelevant passage patterns and six templates for the relevant passages. The new rules are to be implemented in the Stage 4 algorithm to improve tool precision.

## Part-of-Speech Analysis Round II

A second round of part-of-speech analysis also took place within Stage 4. The goal was to identify a natural language processing pattern that goes beyond the search term and template-based methods discovered using manual rule-based classification. In Stage 2, the part-of-speech analysis was used to look for differences in the overall parts of speech used within entire passages. In Stage 4, the analysis was more in depth. Rather than overall passage evaluation, the different relevant passage pattern term sets were examined for their parts of speech. Both the relevant and irrelevant passage classes were inspected, with the hope that certain parts of speech would appear more often within a term set for one class than the other class. For example, if the modification terms appeared as adjectives a majority of the time in the relevant passages and infrequently in the irrelevant passages, then a new pattern could be developed requiring adjective modification terms in order to form a passage. Conversely, part-of-speech types that occur more often in irrelevant passages than relevant passages can be used to create filtering rules.

The part-of-speech study section begins with a description of the software and text used for analysis, followed by the research approach taken. An evaluation of the results is conducted by search term set. The section concludes with a discussion of how part-of-speech may be used for relevant and irrelevant passage classification.

### *Materials*

The Stanford Part-Of-Speech Tagger used in Stage 2 was also used for analysis in this study [86]. The POS tagger assigns the part of speech tags to each word in the document using the Penn Treebank tag set [87]. Following the tag groupings used by

Glier, the 45 Treebank tags are once again categorized into the four major groups of nouns, verbs, adjectives/adverbs, and conjunctions/prepositions [84].

The passages used for the study consisted of the relevant and irrelevant passages from the Mechanical Behavior special issue for the property of strength. A total of 58 relevant and 86 irrelevant passages were inspected.

*Research Approach*

To prepare the passages for analysis, the relevant and irrelevant passage classes was formatted into a text file. The revised Stage 1's algorithm removes special characters so periods were added after each sentence to help with analysis. Each passage was also numbered in order to track the results. The two text files were evaluated using the Stanford POS Tagger. Then, the relevant passage pattern search terms were identified within the results, along with their part-of-speech tag. The number of uses of each part-of-speech tag for each search term set was then tallied and normalized by the number of passages in the class, and then used to search for patterns. The four sets of search terms were the modification terms, causation terms, correlation terms, and comparison terms. The causation and comparison terms could appear anywhere within the passage, as per the relevant passage templates. The modification and correlation terms were only counted if they appeared within the property sentence. An initial survey of the structural design principle key terms' part-of-speech was also begun.

## *Evaluation and Discussion*

The modification terms appeared 49 times within the property sentence in the relevant passages, and 42 times in the irrelevant passages. The parts of speech usages after normalizing by the number of passages are displayed in Table 8, using the full set of tags from the Penn Treebank. In Table 9, the tags are categorized into the modified tag groupings. For both the relevant and irrelevant passages, a majority of the modification terms were adjectives (JJ and JJR). Therefore, adjective modification terms cannot be used as a passage classification pattern. However, within the relevant passages, nouns (NN) and verbs (VB, VBD, VBG, VBN, and VBZ) were used often as well. While the relevant passages had a greater number of nouns and verb usages than the irrelevant passages, noun and verb modification terms were not a requirement for all of the relevant passages. Thus, the utilization of a noun and verb modification term to filter and remove the irrelevant passages would also remove many relevant passages as well.

Table 8. The modification terms parts of speech for the relevant passage class and irrelevant passage class using the Penn Treebank tagset.

| | JJ | JJR | NN | RBR | VB | VBD | VBG | VBN | VBZ |
|---|---|---|---|---|---|---|---|---|---|
| Relevant | 0.121 | 0.207 | 0.190 | 0.069 | 0.000 | 0.017 | 0.086 | 0.121 | 0.034 |
| Irrelevant | 0.198 | 0.174 | 0.035 | 0.047 | 0.035 | 0.000 | 0.000 | 0.000 | 0.000 |

Table 9. The modification terms parts of speech for the relevant passage class and irrelevant passage class using the modified tag groupings.

| | Nouns | Verbs | Adj/Adv | Conj/Prep |
|---|---|---|---|---|
| Relevant | 0.190 | 0.259 | 0.397 | 0.000 |
| Irrelevant | 0.035 | 0.035 | 0.419 | 0.000 |

The causation terms were used 144 times in the relevant passages and 210 times in the irrelevant passages. When normalizing the counts by the number of passages, a causation term is used an average of 2.48 times per relevant passage and 2.44 times per irrelevant passage. The full table containing the 29 different tags for the causation terms from the Penn Treebank set of tags can be found in Appendix G. For both classes, the most often used tags were IN, NN, and WDT. A summary of the relevant and irrelevant causation terms' parts of speech can be found in Table 10. Both classes had a large number of conjunction and preposition causation terms, as well as nouns. The summarized table also includes phrases (such as JJ TO, NN IN, VB TO, and VBZ IN), to (TO), and Wh-determiners (WDT), which were not a part of the major modified tag groupings. Since the causation terms' parts of speech were similar between the relevant and irrelevant passage classes, no patterns could be used for passage classification.

Table 10. The causation terms parts of speech for the relevant passage class and irrelevant passage class using the modified tag groupings.

| | Nouns | Verbs | Adj/Adv | Conj/Prep | Phrases | To | Wh-determiner |
|---|---|---|---|---|---|---|---|
| Relevant | 0.310 | 0.207 | 0.103 | 1.086 | 0.483 | 0.000 | 0.293 |
| Irrelevant | 0.267 | 0.174 | 0.128 | 1.267 | 0.337 | 0.023 | 0.244 |

The correlation term appeared less often within passages, partially due to the requirement that the term appears in the property term sentence. In the relevant passages, the correlation term appeared 17 times, in comparison to the 4 occurrences in the irrelevant passages. The only significant difference between the two classes was that the correlation terms were nouns more often for the relevant passages than the irrelevant passages (Table 11 for the modified groupings, Appendix G for the full tagset). However, the correlation terms also appeared as other parts of speech for the relevant passages. Similar to the modification terms, no pattern can be created to filter out the irrelevant passages without losing relevant passages as well.

Table 11. The correlation terms parts of speech for the relevant passage class and irrelevant passage class using the modified tag groupings.

| | Nouns | Verbs | Adj/Adv | Con/Prep | Phrases |
|---|---|---|---|---|---|
| Relevant | 0.172 | 0.017 | 0.000 | 0.000 | 0.103 |
| Irrelevant | 0.000 | 0.023 | 0.000 | 0.000 | 0.023 |

The final term set are the comparison terms, which were mostly used as propositions or subordinating conjunctions (Table 12, Appendix G). The 31 relevant uses and 42 irrelevant uses were often tagged as “IN”, which includes the comparison terms “than”, “while”, and “whereas”. Once again, the high degree of similarity between the two passage classes caused difficulty in identifying patterns to classify future passages.

Table 12. The comparison terms parts of speech for the relevant passage class and irrelevant passage class using the modified tag groupings.

| | Conj/Prep | Phrases |
|---|---|---|
| Relevant | 0.517 | 0.017 |
| Irrelevant | 0.395 | 0.093 |

An initial evaluation of the structural design principle parts of speech was begun. Nearly all of the principles and principle phrases contained a noun. This pattern may prove useful in future research by enabling the extraction of principle key terms with noun phrase searches.

## *Conclusion*

The different term sets used by the relevant passage templates were analyzed by their parts of speech in an attempt to identify new useful strategies for classifying the passages. While there were some differences in the parts of speech used for some of the term sets, the dissimilarities could not be used to generate patterns for the Stage 4 algorithm. One method to create a unique classification rule would be to have a part-of-speech grouping appear far more often in the irrelevant passages than the relevant passages. The frequently used parts of speech could then be used as a filter to remove the irrelevant passages. However, there were no significant patterns in the irrelevant passages that were not found in the relevant passages as well. The other option to create a new algorithm rule would be to use the part-of-speech groupings that occur more frequently for the relevant passages than the irrelevant passages. Although this was the case for certain term sets, the pattern could not be used to create a filter since the relevant passage pattern terms made use of other parts of speech as well.

## Passage Packet Activity

A passage packet activity was used to identify additional patterns within the two passage classes. After having worked closely with the passages for a considerable amount of time, a check was deemed necessary to ensure that obvious solutions were not being missed. Four graduate-level mechanical engineering design students, with at least an undergraduate level's basic understanding of materials, were given packets with a randomized set of passages to determine their own classification methods. This chapter section covers the development of the packet, and describes the activity and instructions to the students in further detail. It then examines their responses and discusses the qualitative results.

*Materials*

The passage packet contained 49 passages from the Advanced Materials, Materials Science and Engineering, and Mechanical Behavior of Biomedical Materials special issues used in the first three stages of the research. Ten passages were taken from the manually identified Mechanical Behavior passages, four were taken from the issue's revised Stage 1 results, and six passages were taken from the Stage 3 results. The twenty Mechanical Behavior passages were looked over to ensure that none of the passages overlapped and that at least one passage appeared per property. For the other two journal issues, the Stage 1 passages were selected to include at least one relevant passage and one irrelevant passage per property. The passages chosen were randomized based on a random sequence generator. From the Advanced Materials issue, seven relevant passages were selected and ten irrelevant passages were selected. Six relevant and six irrelevant passages were selected from the Materials Science and Engineering issue. All of the passages were three sentences in length.

After randomizing the passage order, a solution packet was created to note the passages' gold standards based on their classification. The irrelevant passages were given a second investigation to determine if they discussed a structure-property relationship for another property. For example, the following passage was considered irrelevant when it was selected for the property of hardness, but the passage actually does discuss a structure and its relationship to toughness.

> *"The bone plates have a circulatory blood system that attaches to the animal through veins. This unusual structure is similar to LAST tiles used on tactical vehicles fig 2b which are hard ceramic hexagonal*

*blocks covered by a polymer. This indicates that the armadillo shell should have excellent impact resistant properties." [85]*

The passage's gold standard needed to be updated to belong to the relevant passage class. The packet contained a total of 23 relevant passages and 26 irrelevant passages.

An introductory page was added to the front of the packet to explain the overall research goal and why the passages required classification. Instructions were printed to inform the participants that they must read through the passages, mark each passage as "good" or "bad", and note the property or properties that would be affected if the passage was "good". An example of a "good passage" and a "bad passage" were also included. The final page of the packet asked the participants to "Please note what you were using to evaluate the passage. Also, what patterns did you notice commonly appear in the 'bad' passages as opposed to the 'good' passages?" A full reproduction of the solution packet can be found in Appendix H. The relevant passages have been italicized and their corresponding structural design principles are underlined.

### *Research Approach*

The passage packet was distributed during a lab meeting to the four graduate mechanical engineering students (three female, one male). The research goal and issues were discussed in further detail, and any questions by the participants were answered. The students were given one week to complete the activity on their own time. Following the return of their packets, they were given three follow-up questions:

1. On a scale of 1 to 5, how knowledgeable are you about the field of materials?
2. Did you have any difficulty reading the passages? If so, why?
3. How did you identify the material structures?

The participants' responses on the final page of the packets were reviewed first. Each mentioned pattern was compared to the previously manually identified Stage 4 classification patterns. If any of the patterns were new, an examination of the pattern usage within the packet was completed. Next, the follow-up questions were considered to try to further elucidate the participants' thought processes when scoring the passages. Finally, the scoring of the passages was studied to determine the commonly missed relevant and irrelevant passages.

### *Evaluation and Discussion*

The participants appeared to put forth good effort in identifying the passages that include structural design principles. Their "good" passages were selected by looking for structure-property relationships, rather than just the mention of a material structure. Also, they considered the passages independently. For example, the same structures were mentioned in multiple passages, but the participants only scored the passages as "good" when the property relationship was also included.

For the final question asking for common patterns appearing in the irrelevant and relevant passages, the participants tended to identify patterns for one class or the other. Within the "bad" passages, one of the participants pointed out that "bad passages usually state/give information about experimental procedures", while another stated that "a lot of the bad passages were describing how to test for certain properties". Both of these comments were considered during the Stage 4 manual rule-based classification, which resulted in the irrelevant passage pattern of method terms. In terms of relevant passage patterns, two participants noticed the pattern of "good passages [including] reasoning/causes and effects/because", which was taken into account by the Stage 4 relevant passage causation term set. A third participant pointed that "a lot of the good passages had an adjective, like 'higher', immediately preceding the mechanical property." The indicated adjective is one of the Stage 4 modification terms. An inspection was completed into the modification term location, and the terms were found to not consistently appear directly prior to the property term. Thus, the Stage 4 relevant passage templates that simply required the use of a modification term within the property term sentence were retained. One of the participants also suggested the use of synonyms for the property terms when searching, which is already in use with the property term sets developed in Stage 1.

One pattern was suggested that was not already planned for the Stage 4 algorithm. One participant suggested using passages that contain the word "structure". The pattern was considered during the manual rule-based classification portion of Stage 4, but was rejected due to the inconsistent usage within the relevant passages. The pattern was revisited by examining the passages in the packet. Other "structure" terms were identified from the passages to include "structures", "structural", "microstructural", "microstructure", and "structured". Only 14 of the 23 relevant passages contained one of these terms, while 6 of the 26 irrelevant passages contained one of the terms. Based on these numbers, the usage of a “structure” term to identify relevant passages would not be useful.

To identify the material structure, the participants mostly relied on their prior materials knowledge. One stated that they identified a material structure “if there was a description about the general makeup of the material (i.e. grain size, shape, etc.)”, while another suggested looking for “structure” terms in the passages before realizing that they sorted passages as good that did not contain a “structure” term. Based on their responses, there was not a semantic-based pattern that could be used to identify the material structures. Instead, it is necessary to draw upon materials domain knowledge.

On average, the participants misclassified 7 of the relevant passages. One of the participants missed 13 of the passages. However, there was only one relevant passage (Passage 13) consistently missed by the participants, which described fiber planes’ orientation effect on toughness. Three of the participants missed Passage 31 which implied the structure-property relationship, but did not directly state it. The other missed relevant passages varied among the participants.

The participants misclassified 3.5 of the irrelevant passages, on average. The most misclassified passage was missed by three participants. The passage stated that there were “structural dissimilarities”, but did not actually describe the structures. Many of the other errors were due to lack of materials knowledge, such as the definition of a gel.

Three of the participants gave themselves a 3/5 when asked to assess their materials knowledge, and the other gave herself a 2/5. The main complaint in terms of difficulty reading the passages came from their unfamiliarity with the materials language and terms used.

### *Conclusion*

During the activity, the participants detected patterns in the relevant and irrelevant passages. The detected useful patterns were already a part of the planned Stage 4 algorithm. While the passage packet activity did not offer new information in terms of pattern identification, it did highlight that the research problem is a difficult one. Even with basic materials knowledge, the participants struggled to identify the relevant passages and patterns of the passages. Text classification is difficult, and this particular application is made trickier with the requirement of materials domain expertise.

## Updated Algorithm

The Stage 4 algorithm was created by updating the Stage 3 algorithm with the patterns identified by the second round of manual rule-based classification (Figure 14). The green shapes represent the steps required for the relevant passage patterns, and the orange shapes are for the irrelevant passage patterns.

Like the previous iterations of the algorithm, the Stage 4 algorithm begins by searching for a biological term to ensure that the paper is describing a biological or bioinspired material. The algorithm then searches for the term sets required for the relevant passage patterns. The algorithm identifies the sentence location of the property terms and the stand-alone terms, and then completes a check to ensure that they are not a part of a property-specific exclusion term phrase. The algorithm also searches for the modification, correlation, comparison, and causation terms. The first irrelevant passage pattern takes place with the search of the term "not" followed by a relationship term. Since the causation terms are similar to the relationship terms in that they describe a causal relationship, the check is made that the word "not" does not precede the causation term.

Next, the term sets are used to identify passages that follow the relevant passage pattern templates. The first template includes the usage of a property term with a modification term in the same sentence, and a causation term within one sentence. A second template contains the stand-alone property term with a causation term within one sentence in distance. Third is the property term with the correlation term in the same sentence. Fourth is the usage of both a stand-alone property term and a correlation term within the same sentence. The fifth template includes the modification and property term within the same sentence, and a comparison term in the same sentence or one sentence away. The final template has a stand-alone term with a comparison term within one sentence. If any of these templates are followed, a three-sentence passage is produced, and then an assessment for the remaining irrelevant passage patterns is completed.

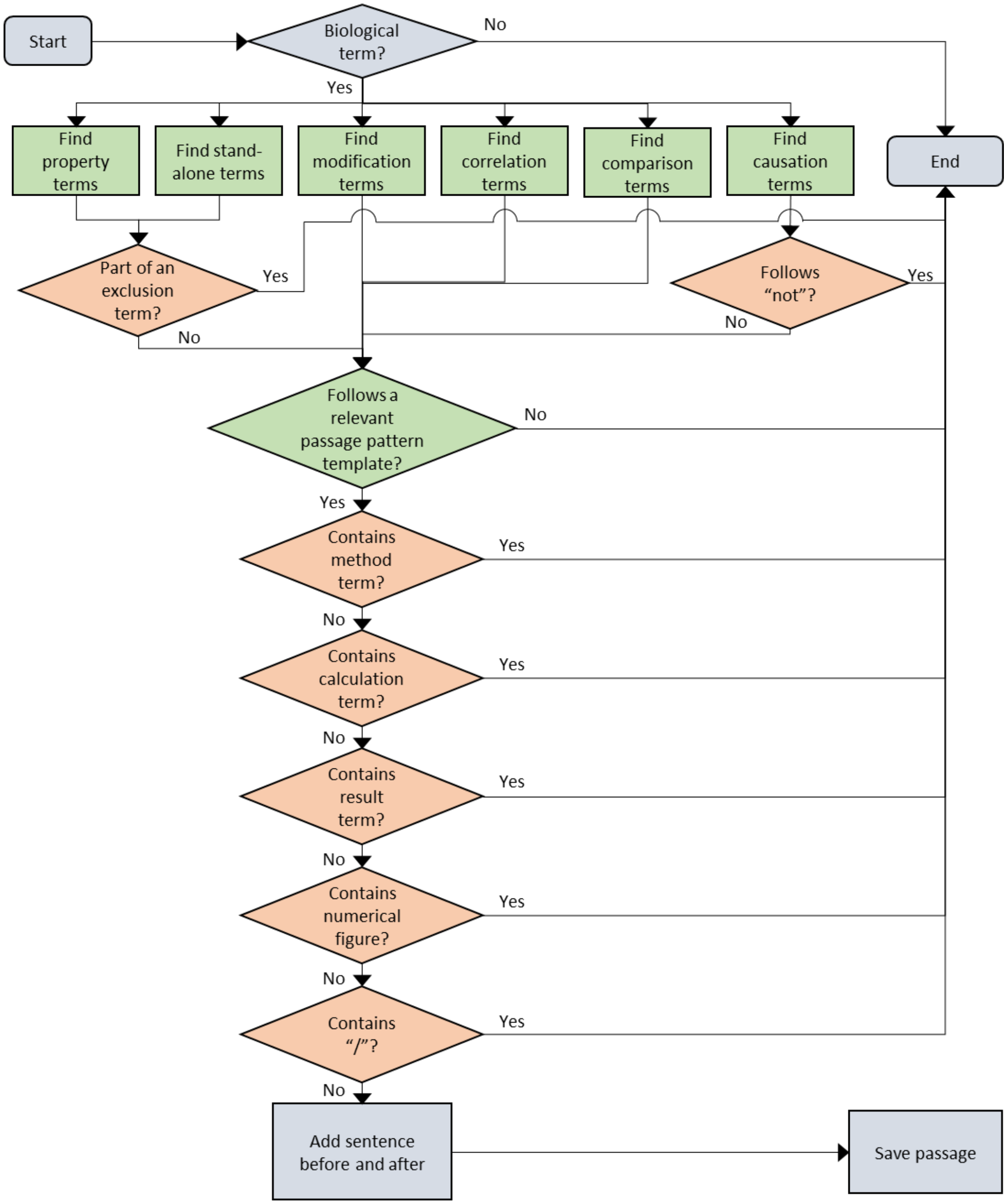



Figure 14. An outline of the Stage 4 algorithm.

Within the passage, a search for the method terms takes place. If a method term appears, then the passage is discarded. If there is not a method term, then a check for a calculation term takes place within the property sentence. If there is not a calculation term, then the algorithm searches for a results term within the full passage. If the passage passes, then a check for a numerical figure in the middle sentence or the use of a forward slash in the passage takes place. If the passage does not contain any of the irrelevant passage patterns, then it is saved.

The full Java program code based on the Stage 4 algorithm can be found in Appendix I. The exceptions to the different patterns, as described in Section 7.2's passage patterns, are included in the code.

## Algorithm Evaluation

A full evaluation of the Stage 4 algorithm was completed. A new set of articles was run through both the Stage 4 algorithm and the revised Stage 1 algorithm with the updated passage length, as used in Stage 3. The output passages were classified as relevant or irrelevant, depending on if the passage contained the expected structural design principles for the article and property. The six individual properties' passages were evaluated, similar to the Stage 3 evaluation. The property combinations would retrieve passages that overlap between the individual properties' results.

The section begins with a discussion of how the articles were selected for testing. Next, a quantitative and qualitative evaluation of the Stage 1 and Stage 4 results takes place.

### *Articles for Evaluation*

A final set of articles was required for the Stage 4 algorithm evaluation. It was difficult to identify another biological materials special issue, so instead 20 biological materials articles were to be chosen from a materials journal. The journals from which the highly cited biological and bioinspired materials articles used in the Stage 1 algorithm development were considered, and the Journal of the Royal Society Interface was selected. The monthly journal occasionally publishes biological materials articles within the standard issues.

Starting from the June 2015 issue, the abstracts were read to select articles. The articles describing biological material structure, rather than material design and development, were desired. Articles regarding biomedical materials and other properties (sensing, adhesion, self-healing, thermal) were avoided. While the algorithm is expected to select the proper articles during use, it was necessary to choose relevant articles for testing in order to evaluate the algorithm's ability to identify structural design principles. To find 20 suitable articles [4, 96-114], it was necessary to go through 41 journal issues to the January 2012 issue.

The articles were read and the structure-property relationship principles were identified manually. The article texts were then formatted into a text file for testing. The same format used from Stage 1 was used here. Each article is represented by a "##", followed by the article information, then a "#", followed by the article text. By use of this format, the article information can be returned alongside the passages so a materials designer can easily reference the article, if necessary.

*Quantitative Comparison*

The revised Stage 1 and Stage 4 recall and precision scores can be found in Table 13 and Table 14, respectively. Comparing the two stages, the overall percentage of useful passages improved from 46% to 65% with the final updated algorithm. As expected, the precision improvement was at the expense of recall. The revised Stage 1 algorithm was able to identify most of the principles (94% recall), while the Stage 4 algorithm achieved a recall of 63%. However, the Stage 4 algorithm had the advantage of returning 796 fewer passages than the revised Stage 1 algorithm.

Table 13. An evaluation of the recall for the revised Stage 1 and Stage 4 algorithms.

| | Stage 1 (Revised) | | | Stage 4 | | |
|---|---|---|---|---|---|---|
| Properties | Identified Principles | Total Principles | Recall | Identified Principles | Total Principles | Recall |
| Density | 13 | 13 | 100% | 6 | 13 | 46% |
| Ductility | 8 | 9 | 89% | 6 | 9 | 67% |
| Hardness | 19 | 21 | 91% | 10 | 21 | 48% |
| Stiffness | 63 | 66 | 96% | 42 | 66 | 64% |
| Strength | 28 | 30 | 93% | 17 | 30 | 57% |
| Toughness | 31 | 33 | 94% | 28 | 33 | 85% |
| **Total** | **162** | **172** | **94%** | **109** | **172** | **63%** |

Table 14. An evaluation of the precision for the revised Stage 1 and Stage 4 algorithms.

| | Stage 1 (Revised) | | | Stage 4 | | |
|---|---|---|---|---|---|---|
| Properties | Relevant Passages | Irrelevant Passages | Precision | Relevant Passages | Irrelevant Passages | Precision |
| Density | 47 | 64 | 42% | 5 | 11 | 31% |
| Ductility | 14 | 58 | 19% | 5 | 28 | 15% |
| Hardness | 62 | 100 | 38% | 15 | 6 | 71% |
| Stiffness | 199 | 165 | 55% | 74 | 26 | 74% |
| Strength | 110 | 132 | 46% | 41 | 17 | 71% |
| Toughness | 61 | 60 | 50% | 38 | 10 | 79% |
| **Total** | **493** | **579** | **46%** | **178** | **98** | **65%** |

To compare the overall ability and accuracy of the two algorithms, the traditional F-measure was utilized. The F-measure is often used for information retrieval and document classification, and is the harmonic mean of recall and precision (Equation 3).

$$F = \frac{2*precision*recall}{precision+recall} \tag{3}$$

In Stage 3, the revised Stage 1 algorithm, which returns three-sentence passages and requires that the property and relationship term appear in the same sentence, achieved an F-measure of 0.589. The revised Stage 1 algorithm slightly outperformed the Stage 3 algorithm's 0.576 F-measure.

For the Stage 4 algorithm evaluation, the revised Stage 1 algorithm's F-measure was 0.618. The increase in the Stage 1 algorithm's F-measure was due to the article set. The recall from the previous evaluation to the current evaluation improved from 82% to

94%, suggesting that structural design principle descriptions followed the Stage 1 pattern more closely in the Royal Society Interface articles. However, the precision of the revised Stage 1 algorithm was 46% for both evaluations. The similar precision values confirm that the increase in precision from the original Stage 1 algorithm to the revised Stage 1 algorithm was not a fluke during the previous chapter's evaluation. The reduced three-sentence passage length, and the associated relationship and property term requirement, does consistently improve the Stage 1 precision score over the original Stage 1 algorithm.

The Stage 4 algorithms' F-measure was 0.639, which was greater than the revised Stage 1 algorithm's score. The algorithm's ability is improved, even with the significant reduction in number of passages returned. For the six material properties, the revised Stage 1 algorithm returned 1,072 passages, which consisted of 110,779 words, while the Stage 4 algorithm returned only 276 passages containing 27,444 words. The Stage 4 algorithm saw a 74% decrease in the number of passages and a 75% decrease in the number of words. Fewer passages and words translates to less reading required by the user to identify potential structural design principles.

The revised Stage 1's 0.618 and the Stage 4's 0.639 F-measure values were similar to the new work within the biomedical text classification field presented in Chapter 2. McKnight and Srinivasan's best method for the categorization of sentences in medical abstracts into introduction, method, result, and conclusion classes achieved a F-measure of 0.613 [67]. Similarly, Dobrokhotov et al.'s classification of article titles and abstracts for database curation obtained a F-measure of 0.636 [66].

A total of 172 principles were identified within the Royal Society Interface articles. Both algorithms identified 104 of the principles, and both missed 5 of the principles. The revised Stage 1 algorithm's improved recall identified 58 principles that the Stage 4 algorithm overlooked. However, the Stage 4 algorithm was able to identify 5 principles that the revised Stage 1 algorithm missed. Based on these results, the algorithms are running separate searches and one is not merely a subset of the other. If a materials designer wants to identify as many principles as possible, they may want to run both stages of the algorithm.

Looking more closely at the precision and recall by property, the overall trends are not displayed by all of the individual properties. Upon examination of the recall values, density and hardness had noticeably lower recall than the other properties for the Stage 4 algorithm. Density and hardness also had the fewest number of returned passages from the final algorithm. It is likely that the restrictions that lead to the fewer number of passages is also removing the relevant passages as well. Therefore, if the program user uses the Stage 4 algorithm and the tool returns fewer passages than expected, the user may want to rerun the articles through the revised Stage 1 algorithm to widen the search. The Stage 1 algorithm had high recall for both density and hardness. Density and ductility also had better precision for the revised Stage 1 algorithm. Principles are being identified with the Stage 4 algorithm, but a greater percentage of irrelevant passages are found. One of the reasons for the low precision for those properties by the Stage 4 algorithm is that there are fewer passages so the irrelevant passages have a greater effect. One of the articles spent much of the paper discussing extensible threads, but did not discuss the thread structure, which resulted in 22 of the 28 irrelevant ductility passages.

The revised Stage 1 program had a reduced run time compared to the Stage 4 program. For the 20 Royal Society Interface articles, all combinations of the six material properties, which is a total of 63 combinations. The revised Stage 1 program took 1 minute and 13 seconds, and the Stage 4 program took 1 minute and 53 seconds. While the Stage 1 program was slightly faster to run, the results from the Stage 4 program would be faster for the user to read. Overall, the revised Stage 1 algorithm surpassed the Stage 4 algorithm in terms of the number of principles identified and program runtime, but the Stage 4 algorithm outperformed in terms of precision and reducing the number of passages.

*Qualitative Comparison*

During passage evaluation, it was noted that the Stage 4 tool's passage readability was much improved over the revised Stage 1 tool. The Stage 1 algorithm returned more procedural and method descriptions that were not output by the Stage 4 tool. An example of such a passage for the property of hardness is:

> *"in this work the skeletons of two coral species solitary balanophyllia europaea and colonial stylophora pistillata were investigated by nanoindentation*
>
> *the hardness hit and youngs modulus eit were determined from the analysis of several loaddepth data on two perpendicular sections of the skeletons longitudinal parallel to the main growth axis and transverse*

> *within the experimental and statistical uncertainty the average values of the mechanical parameters are independent on the sections orientation" [108]*

The Stage 4 tool generally produced fewer passages per article, and the passages focused more on the structural design principles of the article. The fewer passages may cause the tool's user to read through the passages more carefully. With the addition of fewer irrelevant passages within the Stage 4 results, the program's output appears more valuable compared to the revised Stage 1's results.

Another difference between the two tools is that the revised Stage 1 version returns more results for articles that do not have structural design principles for the property in question. The Stage 1 tool returns more mentions of the property, but in terms of procedure or general results, rather than descriptions relating the property to a structure. The Stage 4 algorithm worked to remove these forms of the irrelevant passages with the number, ratio, method, calculations, and results patterns. The resulting differences can easily be seen in the program output for the property of hardness from the Stage 4 tool (Appendix J) and the revised Stage 1 tool (Appendix K). Also evident in the program output is the large difference in the number of returned passages.

It may be possible to update the Stage 4 tool to achieve a range of precision and recall values, rather than just focusing on precision. The six irrelevant passage patterns added to the Stage 4 tool are easy to switch on and off. By disabling patterns, the filtering restrictions are eased and more passages can be output, which would likely lead to an increased recall. The user can select which patterns to use based on the recall-precision trade-off they desire. However, even with all of the irrelevant passage patterns disabled, there will still likely be fewer passages returned with a lower recall than the revised Stage 1 algorithm. The relevant passage patterns for the Stage 4 tool are stricter than the algorithm used by the Stage 1 tool.

## Conclusion

One common application for manual rule-based classification is to divide newswire articles into classes [59]. For example, Jacobs and Rau identified articles that were about a business takeover by using keyword filtering from a specified set of keywords such as “merger” and “acquisition” [115]. While the two businesses were not known by the program prior to classification, the relationship between the two was known. This dissertation work is difficult because while a property is known, the structure and the relationship are not. The program can search for the selected material property with property keywords, but the property terms are used often throughout materials articles without reference to material structures, leading to a high number of irrelevant passages. This chapter describes the work completed to decrease the number of irrelevant passages output from the program algorithm.

The chapter began with an introduction to the prior algorithms. At the end of Stage 3, an algorithm with a recall of 80% and an improved precision of 45% was developed. For the final Stage 4 algorithm development, the goal was to create a tool that would require less user effort in identifying principles by further increasing precision.

Three approaches were examined in the chapter. The first was a more strict manual rule-based classification system. Six irrelevant passage patterns and six relevant passage templates were identified. The second approach was a more in-depth examination of parts of speech for the different passage sets utilized by the relevant passage patterns. However, no significant patterns in the relevant or irrelevant passages were produced that could be used as useful strategies for filtering out irrelevant passages. The final approach was to have others also look for patterns within the relevant and irrelevant passages. A passage packet activity was completed by four graduate mechanical engineering students. While no new rules for classification were discovered, the problem difficulty and need for domain expertise was emphasized.

The Stage 4 algorithm was created and evaluated against the revised Stage 1 algorithm. The rules and templates manually identified from the Stage 3 passages were successfully applied to a new set of articles. As expected, the Stage 4 algorithm had an improved overall precision (65%), although the recall was reduced to 63%. Along with the improved precision, the Stage 4 algorithm resulted in a substantial reduction in retrieved passages, which would make the process of identifying potential principles significantly faster for a materials designer. The precision improvement and reduction in retrieved passages was likely due in large part to the irrelevant passage patterns, which act as filters. Through the application of these patterns, many passages describing methods, calculations, equations, and basic results were removed. Passage readability was also improved with the Stage 4 algorithm. The revised Stage 1 algorithm was simply attempting to show that a relationship exists within the passage, while the Stage 4 algorithm aims to retrieve passages that describe how the property changes and the structural cause of the change. The Stage 4 relevant passage templates are stricter and require more from the passage, which leads to more valuable results.

The two algorithms can be used in different ways by the user. If a materials designer wants to quickly survey some structural design principle options for their property of interest, they can use the Stage 4 algorithm to retrieve fewer passages with higher precision. However, if the user wants a deeper search with more information and sources of inspiration, and is willing to invest more time, the Stage 1 algorithm would be recommended due to its high recall. A materials designer can identify principles more quickly with either algorithm as compared to reading all of the articles on their own. Both algorithms are returning structural design principles, which can then be used for the design of bioinspired materials.

# CHAPTER VII

# IDENTIFIED PRINCIPLES AND VALIDATION

An algorithm and text mining tool were developed that return passages describing biological materials' structural design principles to the program user. The evaluation of each stage of the algorithm was completed during that stage's algorithm development. The closing task is to analyze and validate the results from the final iteration of the program. The goal of this chapter is to demonstrate that bioinspired materials can be created based on the design principles, and that the principle structures can indeed improve material properties. Since there is not enough time to develop and design the materials, previously developed bioinspired materials are examined. The most widely researched biological material principles are the ones most likely to have reached the materials design domain. Therefore the ten most referenced structural design principles are identified using the Stage 4 tool, and a search for bioinspired materials based on these principles is conducted. Bioinspired materials designed from the principles are assessed to determine if the utilization of the principles result in the expected property improvements.

The chapter begins with a description of the research approach taken to select the principles and to identify the bioinspired materials. Next, each of the ten structural design principles are discussed in detail along with the biological materials making use of the principle. A report of how the principle was utilized for a bioinspired material follows, as well as an examination into the impact of the principle on the bioinspired material’s properties.  The chapter concludes with a discussion of the text mining tool’s ability to identify principles and the utilization of the structural design principles in material development.

## Principle Identification Research Approach

For the tool validation and empirical product study, structural design principles and relevant bioinspired materials were selected for investigation.  The principles were identified from the four article sets used during algorithm development.  There were a total of 72 articles from the Advanced Materials Biological and Biomimetic special issue [73], Materials Science and Engineering C special issue [80], Journal of the Mechanical Behavior of Biomedical Materials special issue [79], and Journal of the Mechanical Behavior of Biomedical Materials.

First, the principle-property spreadsheets created for each of the four article sets during algorithm development were consolidated in order to organize all of the principles.  Different wordings were used when first identifying the principles, leading to overlap between the principles.  For example, the “unravel/unfold” principle and the usage of “sacrificial domains” principle were combined because the domains are sacrificed when unfolded or unraveled.  Another example of principle organization was based on the identified principles of “fiber composite” and “fiber orientation”.   The “fiber orientation” principle was chosen over “fiber composite” because it is a more specific description of the material structure.

The four article sets were then run through the Stage 4 algorithm. Each of the returned passages was examined to determine the principles noted within each article. A count of the number of times each principle appeared was conducted. The top ten most-referenced principles within the 72 articles were selected for further analysis. For each of the selected principles, the articles that discussed the principles were further examined to identify the biological materials that made use of the principle and the principle's specific effects on properties. This process is similar to what may be used by a materials designer. The overall tool's objective is to inspire new ideas for material development, and the program is expected to return passages containing structures that will affect the property of interest. However, the text mining tool does not return all of the information necessary to fully design a new material. The program output includes the article information corresponding to the passages so that the tool's user can easily find the article to obtain more detailed information on the material's structuring and effects.

A search was made using Google Scholar with the term "bioinspired" and the principle key terms to identify materials created based on the principles. The bioinspired materials articles were read to understand how the principles were applied to the new material and how the material properties were affected. A comparison of the property effects of the bioinspired materials to those of the biological materials was completed. Often, the principle structures were related to one or two key properties, although other properties were mentioned on a case-by-case basis.

## Identified Principles and Related Bioinspired Materials

In this section, each of the ten principles are examined at length. The principles are introduced in the order they were most referenced. Each principle discussion includes the principle usage in both the biological materials and the bioinspired materials. Within a structure, there may be many factors and considerations that can each affect the material properties. For example, nacre's brick-and-mortar structure can inspire bioinspired materials based on the geometry of the platelets [16], the nanoscale asperities on the platelets [116], or the rotation between platelet layers [117]. Individual bioinspired materials do not contain all of these factors as they draw inspiration from the biological material rather than a complete replication of the natural material. An examination of some of the applied factors are conducted for the selected principles.

### *Porous Architecture*

The most commonly mentioned structure-property relationship in the articles was the usage of porosity and its effect on stiffness. Porosity's relationship to density was also regularly noted, with other properties discussed on a case-by-case basis. The usage of pores can lead to a stiff, yet lightweight material. Porosity was found to affect stiffness in a variety of biological materials such as squid sucker rings [24], osteonal cortical bone [93], trabecular bone [24], wood [24], sea urchin spines [108], marine worm housing tubes [111], red-bellied woodpecker beaks [4], and silkworm cocoons [98]. Density's relationship to porosity was discussed for avian feathers [5], the red-bellied woodpecker beak [4], and silkworm cocoons [98]. The squid sucker rings' porosity was also mentioned as affecting toughness "by introducing a potential crack-arresting mechanism at the boundaries between the two constituent materials" [24].

A porous material was created using bone as an inspiration [118]. A glass scaffold with 60% porosity was produced using a direct-ink-write layer-by-layer assembly technique with a hydrogel-based glass ink. The scaffold was three-dimensional, and required precise filament diameters, spacing, and number of layers in order to produce the pores and struts required for high strengths. An SEM image of the surface of the bioinspired material can be found in Figure 15. The low-density bone-inspired material successfully achieved an elastic modulus and average strength similar to those of human bone. While the work described in this research was still in the beginning stage, the authors proposed that "the work in this study might provide a new avenue for the fabrication of light, yet strong materials" [118]. By taking inspiration from a biological material's principle, the expected resulting properties of the bioinspired material were positively affected.

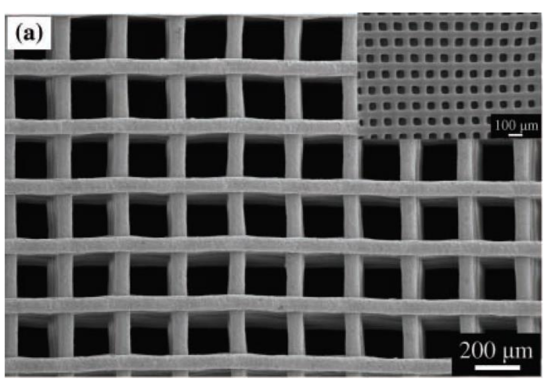



Figure 15. SEM image of a highly porous material inspired by bone [118].

### *Layered Structure*

Many biological materials make use of a layered structure, such as abalone shells [5], sponge spicules [119], armored fish teeth [92], bone [93], and fish scales [103]. The layered structure is able to provide strength, toughness, or stiffness, although generally one or two of the properties are enhanced rather than a combination of all three. There are various types of layered structures within the overarching principle. One such structure is the crossed-lamellar structure, which often corresponds to a high toughness. Materials containing a crossed-lamellar structure include shell structures [119] and river fish scales [5]. Other layered structures include the plywood structure, as found in bamboo [120], or the twisted plywood structure, as can be seen within the dorsal part of a spider's cuticular pad [101]. The differing layer thicknesses also influence the material properties, especially the material stiffness, as observed in tooth [92], spider pad [101], snake integument [103], and bird feather [103].

The bioinspired material used for inspection mimics the crossed-lamellar structure that is commonly found in molluscan shells [121]. Among the molluscan shell microarchitectures, the crossed-lamellar microstructure is the most common and has the highest fracture toughness. The biological crossed-lamellar structure contains five different length scales with a small amount of organic material between the lamella of each length scale (Figure 16). The bioinspired material was created from silicon films containing two-length scales, which were then laminated into a three-film micro-composite. A schematic of one of the films can be found in Figure 17. The films had varying arrangements of angled interfaces in order to mimic the lamellae orientations in the molluscan architecture.  Following the deposition of the micro-composite silicon layer, reactive ion etching was used to dig trenches that were filled with a photoresist to simulate the biological organic interfaces.  An SEM image of the resulting bioinspired material is included in Figure 18.  The authors found that the ratio of the energy dissipated by cracking the bioinspired material into two pieces to that of pure silicon was 36:1.  Similar to the biological materials with crossed-lamellar structures found using the text mining tool, significant toughness was displayed in the bioinspired material.

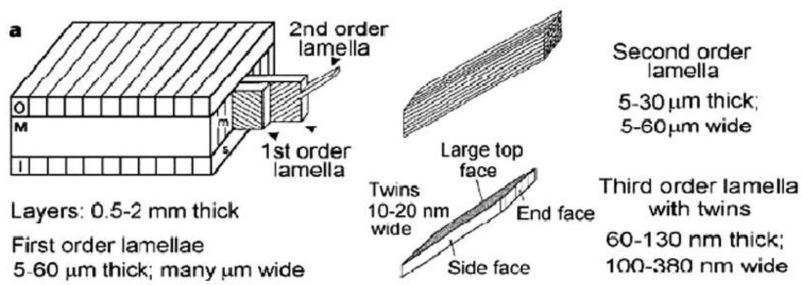



Figure 16.  Crossed-lamellar microstructure of molluscan shell [121].

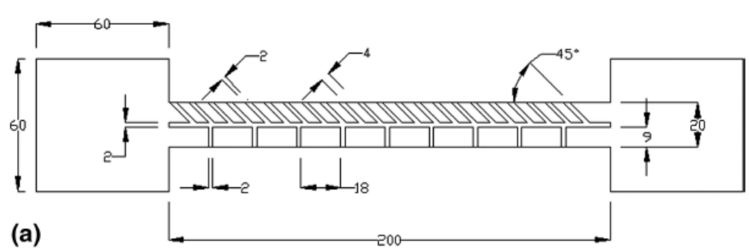



Figure 17.  Schematic of one of the bioinspired two-length scale films.  Dimensions in micrometers [121].

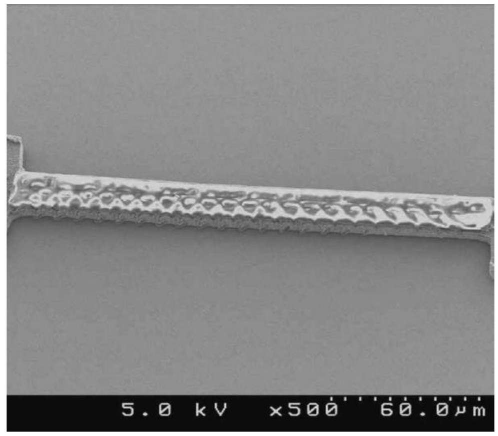



Figure 18. SEM image of a micro-composite inspired by the crossed-lamellar microstructure of molluscan shells [121].

*Fiber Orientation*

The orientation of fibers within a biological fiber composite affects the material's stiffness and strength. An example passage discussing the principle found in the tail feathers of a Toco Toucan is included:

> *"… temporal effects and fiber alignment gradients from the proximal to distal end may contribute to an increase of at least 100% in stiffness or a decrease in failure strength by more than 200%." [95]*

Fiber angles with respect to loading conditions affect a variety of materials. Along with toucan feathers, the wing feathers of swans [95], ostrich feathers [4], and general bird flight feathers [103, 104] have been found to display the fiber orientation principle. Plants, such as spruce and yew wood, have rays and cellulose microfibrils that are helically wound around wood cells, which affect material properties [122]. Palm and bamboo utilize the alignment of cellulose fibrils to improve stiffness along the stem

[114, 120]. The collagen fibers in dentine [123] and the fibers in woodpecker beaks [4] also make use of alignment in order to alter properties.

The chosen bioinspired material was not inspired by one biological material in particular, but rather the idea of the orientation of reinforcements within a composite [124]. There are non-bioinspired engineered fiber composites, but the fibers are often on a larger length scale than those commonly found in nature. For this principle, a bioinspired material was selected that utilizes aligned reinforcements that are on a similar scale to those in biological materials. Composites containing fiber orientation at the proper scale were not found so instead a bioinspired material that makes use of platelets in place of fibers was chosen. The material was created using platelets that were 7.5 μm long and 200 nm thick. The reinforcing platelets were coated so that they would be surface-magnetized, and an ultralow magnetic field was applied to align the platelets (Figure 19). The homogeneously reinforced composite was composed of 20 vol % alumina platelets in a polyurethane matrix. When the reinforcements were oriented parallel to the applied load, the elastic modulus of the composite was 2.8 times higher than that of the matrix material and slightly higher than that of the perpendicular reinforcement material (Figure 20). In terms of strength, the parallel-oriented platelets achieved a 63% higher yield strength than pure polyurethane, and an 85% increased strength over the perpendicularly-aligned platelets. The parallel-oriented composites were also found to have increased hardness over the perpendicular-oriented composites and the pure matrix material. The material properties resulting from reinforcement orientation exhibit a similar trend to biological materials containing aligned fibers.

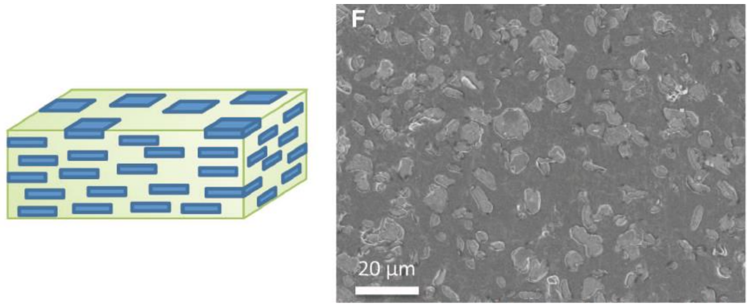



Figure 19. Schematic and SEM image of a composite with in-plane aligned reinforced composites [124].

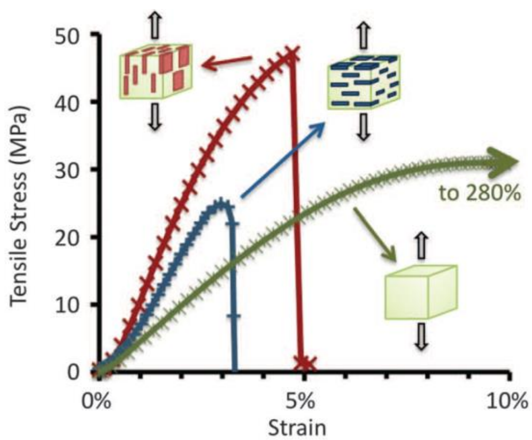



Figure 20. Stress-strain curves of the materials with composites aligned parallel to the load (red), perpendicular to the load (blue), and pure matrix material (green) [124].

*Graded Structure*

Gradients are used in biological materials to primarily influence material stiffness, although density, hardness, strength, and toughness can be affected as well. There are different goals for the graded structure. Many materials use them to achieve a high specific bending stiffness, including bamboo [120], palm [120], and squid sucker rings [24]. Another usage for gradients is to mitigate modulus mismatch, such as for mussel byssal threads [26]. Gradients found in snake skin may have been an adaptation to locomotion and to minimize wear [103]. To achieve a gradient in properties, different methods are used throughout nature. One method is to change the amount of components throughout the material. Armored fish teeth use a gradient of mineral content [92], collagen tissue uses a gradient of collagen content [105], and snake skin uses a gradient of alpha and beta keratin [103]. Squid sucker rings are able to achieve property gradients by varying the porosity in the material [24]. Bamboo and palm have radial gradients by varying the volume fraction of fibers and vascular bundles from the center of the stem to the periphery [100, 120]. Yet another method is to vary the area fraction of the layers in a composite structure, as noted in a study of woodpecker beaks [4].

Bioinspired materials have been created that mimic the complex graded compositions of biological materials and their resulting non-uniform properties. The selected bioinspired material took inspiration from the smooth transition in the spatial distribution of reinforcing elements in bamboo, mollusk shells, teeth, and exoskeletal structures [125]. The bioinspired material made use of MPS-functionalized $Fe_3O_4@SiO_2$ core@shell nanoparticles and a UV-curable hyperbranched polymer (HBP) to act as the matrix to form 150 μm thick films. By applying magnetic field gradients to the superparamagnetic nanoparticles prior to curing, a gradient in the composition was achieved. The composition development was found to be affected by both the viscosity of the nanoparticle suspension, which was reduced by using MPS-functionalized nanoparticles, and the intensity of the magnetic force. An SEM-EDX image of the bioinspired graded material with 8 vol % nanoparticles using two magnets in repulsion can be seen in Figure 21. After UV curing, the elastic modulus and hardness of the materials were measured. The material was found to achieve gradations after one hour. After 24 hours, there was a difference of 70% elastic modulus and 150% hardness between the particle-depleted region and the particle-enriched region, and the gradient was found to be continuous in composition. By varying the amount of the nanoparticle component, the bioinspired material achieved graded properties analogous to some of the biological materials following the graded structure principle. The results are promising for the development of soft materials that require a locally tunable composition.

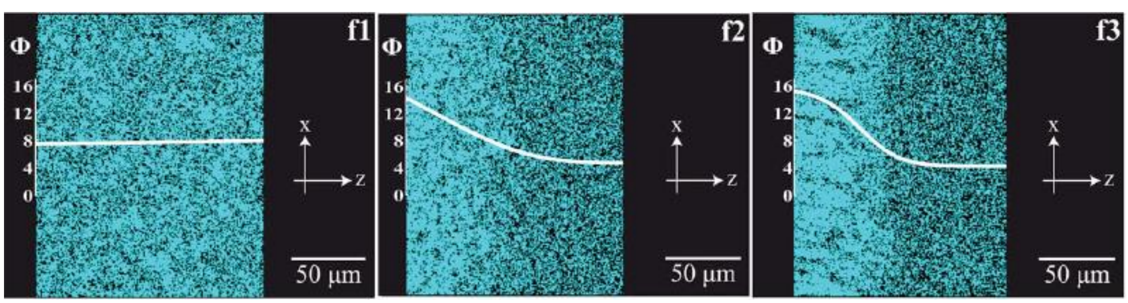



Figure 21. SEM-EDX image of the graded material showing the silicon content with superimposed curves displaying the nanoparticle volume fraction at time [125]: (left) 0 hours, (center) 1 hour, (right) 24 hours.

*Spatial Scale*

As the spatial scale of a biological material component structure decreases, generally the overall material strength increases. At times, this is accompanied with a hardness increase and toughness decrease, which Meyers suggests is due to Weibull statistics [5]. For both spider silk and cocoon silk fibers, a diameter size and gauge length decrease correspond to property improvement due to the lowered likelihood of flaws and flaws that exceed a critical size [110, 126]. Gao et al. noted that in hard biological tissues, the smallest scale in the hierarchical structure is generally on the nanometer scale, such as for teeth, vertebral bone, and shell mineral platelets [14]. They found that the nanometer scale is important to biological materials because materials are insensitive to flaws at that scale. The mineral nanometer particle size delivers optimum strength, which was noted by several of the articles identified by the Stage 4 tool [96, 108].

The bioinspired material drew inspiration from nacre's layered aragonite platelets [127]. The polymer nanocomposite used nanoplatelets at a similar scale and also made use of the layered structure principle. The polymer nanocomposite was composed of hexagonal Gibbsite nanoplatelets with an average diameter of 188 nm and a thickness between 10 and 15 nm. Using electrophoresis, the nanoplatelets were aligned and assembled, as seen in Figure 22. Next, a spin-coating process was used to fill the interstitials between the nanosheets with ethoxylated trimethylolpropane triacrylate (ETPTA) to form the nacre-inspired nanocomposite (Figure 23). The resulting gibbsite-ETPTA nanocomposite had strengths two times higher and modulus three times higher than pure ETPTA polymer.  The increase in strength from the nanometer scale platelets is likely partially due to the flaw insensitivity at that scale, as established by biological materials.

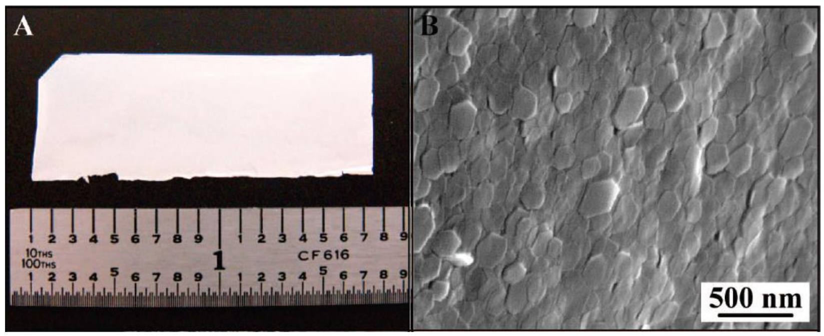



Figure 22.  (Left) Photograph of the freestanding gibbsite film composed of aligned nanoplatelets.  (Right) SEM image of the top view of the gibbsite film sample [127].

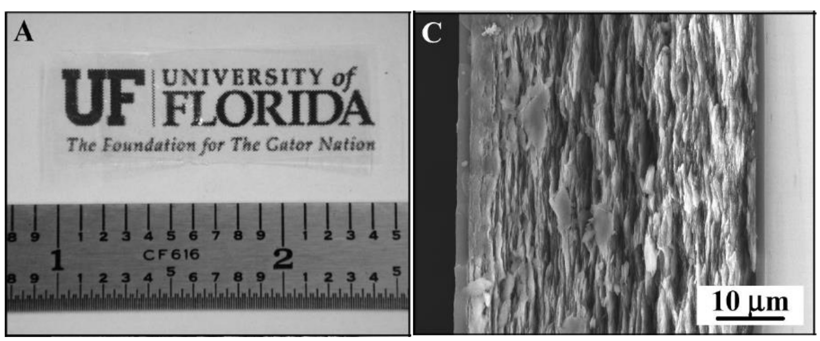



Figure 23.  (Left) Photograph of the freestanding transparent gibbsite-ETPTA nanocomposite film.  (Right) SEM image of the cross-sectional view of the nanocomposite film [127].

*Tubules*

According to the 72 articles used for analysis, the utilization of tubules and hollow vessels within biological materials often has a significant effect on stiffness and density, and occasionally toughness and strength as well. Taylor and Dirks noted that many load-bearing structures in organisms are in the form of a long, hollow tube [113]. They were able to abstract the limb segments of endoskeletons and exoskeletons to fit within that structure for both vertebrates and arthropods. Tubes can be found within a variety of length scales in nature. On the macroscale, a bamboo culm and other tube-shaped plants make use of the principle in order to improve specific flexural rigidity [114, 120]. Bones and other limb segments are also forms of macrostructural tubes [113]. Within hardwood, 0.5 mm diameter vessels are used to transport water. While the main objective of the vessels is not to improve material properties, the vessels do help lower density and increase compliance [123]. In the microscale, tubules are found in dentin [123], vascular bundles consisting of hollow vessels and sclerenchyma cells are in bamboo tissue [100], and a well-developed pore canal system forming a honeycomb-like structure are observed in lobster and crab cuticle [1]. Hollow vessels are even seen in the nanoscale, such as the hollow nanofibrils in the housing tubes of marine worms [111].

A structural materials inspired by tubes in biological materials was not found. However, hollow tubes and tubules were identified in two separate review articles describing bioinspired design strategies. Wegst et al. discussed common biological structural material design motifs, and mentioned that usage of bamboo hollow-tubes causes an improvement in flexural rigidity [2]. The authors also discussed how dentinal tubules in human teeth can improve toughness because microscopic cracks are drawn to the tubules, which result in crack path deflections. McKittrick et al. also noted that tubules were a common theme within biological materials and cited the usage of the structure in bones, teeth, antlers, horns, and hooves [30]. They also noted that the tubular structures can aid in energy absorption. The authors closed with a material design concept using the tubule principle. Since two articles also individually identified the use of hollow tubes as a design strategy for bioinspired materials, it is highly likely that the principle is a good design principle. However, the difficulty in creating a bioinspired material containing tubular structures has hampered principle application. Further technological advances are required to physically create the currently conceptual material designs.

*Brick-and-Mortar Structure*

The brick-and-mortar structure, found in nacre, bone, and teeth, is known for its combination of toughness and strength [16]. For nacre, it has long been thought that the mineral aragonite tablets, acting as the "brick", contribute to the strength of the mineral, while the interlayer organic matrix, the "mortar", provides the toughness by dissipating strain through shearing [2, 96, 119]. There are a variety of features within the structure that also promote material toughness and strength. The thin organic layer further improves toughness by deflecting platelet cracks that enter the matrix [108, 128]. For strength, the platelets' contributing features include the platelet thickness scale and aspect ratio, which is near optimal for strength [108]. To further promote toughness, there are asperities on the platelets [119], mineral bridges between the tiles [5], and an allowance for platelet slip and platelet deformability [108]. The alignment of the components promote toughness as well. The platelets are generally aligned along one direction [108] and the platelets overlap, which aids in blocking crack propagation [4]. The platelets are also interlocked due to the rotation between the platelet layers as they are stacked [117, 119]. There are many features of nacre, bone, and teeth that can be mimicked to create bioinspired materials. However, as Barthelat advises, bioinspired materials should not endeavor to replicate every feature of the biological materials [129].

Since nacre is one of the most widely researched biological materials, there are many bioinspired materials to peruse. Two materials that are both inspired by the brick-and-mortar structure are presented, but the two take very different approaches by mimicking different features.

The first selected bioinspired material was created by Bonderer et al. [16]. They took inspiration from nacre, bone, and teeth's brick-and-mortar structure and created a platelet-reinforced hybrid film. They selected components that had similar properties to nacre's and mimicked the aspect ratio and size of the mineral platelets. The platelet aspect ratio was chosen so that the composite would fail by platelet pull-out rather than platelet fracture. Alumina was used in place of aragonite, and a ductile chitosan was used in lieu of nacre's organic component. The highly aligned brick-and-mortar structure can be seen in Figure 2. The bioinspired layered hybrid films containing 10 vol % alumina and 15 vol % alumina had high strength, ductility, and toughness. The 15 vol % artificial composite was able to achieve a yield strength that was six times greater than that of the pure chitosan matrix material. Compared to the biological inspiration materials, the bioinspired material had a far lower inorganic component. For example, nacre is approximately 95% inorganic. Despite the decreased amount of platelets, the bioinspired materials achieved strengths similar to or greater than the natural inspiration materials (Figure 24). The artificial composites' ductility were substantially greater than the biological materials, which contributed to the order of magnitude increase in toughness of the hybrid film. The bioinspired material was able to take from the biological brick-and-mortar principle in order to create a new material with toughness and strength beyond those of the inspiring materials.

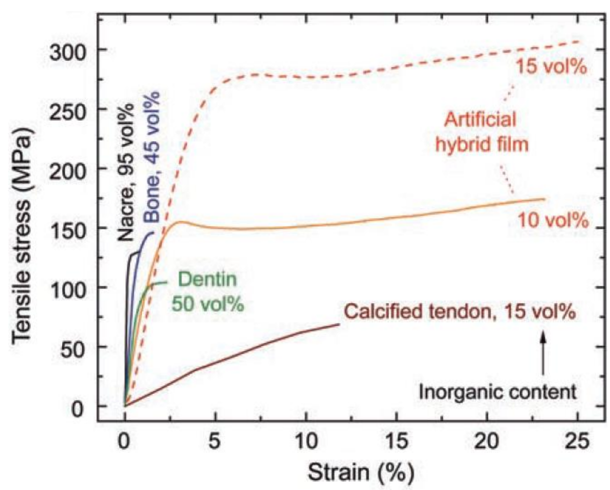



Figure 24. Stress-strain curves of the artificial hybrid film at 10 vol % and 15 vol % platelets (red) compared to biological material [16].

The second bioinspired material also used nacre's brick-and-mortar structures as a source of inspiration [129]. But rather than looking at the platelet size and ratio, the second material drew inspiration from tablet sliding and the mineral bridges. Nacre, bone, and tendon all allow for platelets to slide on one another when in tension. The local sliding expansions provide high strains at the macroscale. The increased material ductility then contributes to increases in toughness. Nacre has 'dovetail' tablets, which are thicker at the ends than the center (Figure 25). The tablet's resulting wavy geometry causes progressive locking at the tablet interfaces during the sliding process, which increases sliding resistance. The bioinspired material uses PMMA for the tablets, which also have dovetail ends. Figure 26 shows the macroscale dimensions of the tablets. There is no 'mortar' component within the material. Rather, mineral bridges are used to hold the tablets in place by means of brass fasteners in the tablet core regions (Figure 27). As anticipated, the artificial nacre assembly did have significantly improved ductility over the pure tablet material PMMA. The failure strain increased from 1-2% to nearly 10% by using the brick-and-mortar structure as inspiration. The bioinspired material was made to be more ductile, even without the use of a replacement for the organic component. The improved ductility would in turn improve the material's overall toughness.

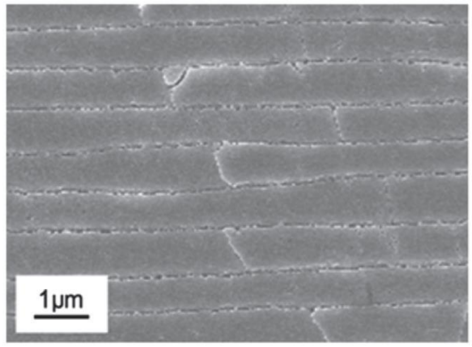



Figure 25. SEM image of nacre's platelets with 'dovetail' ends [129].

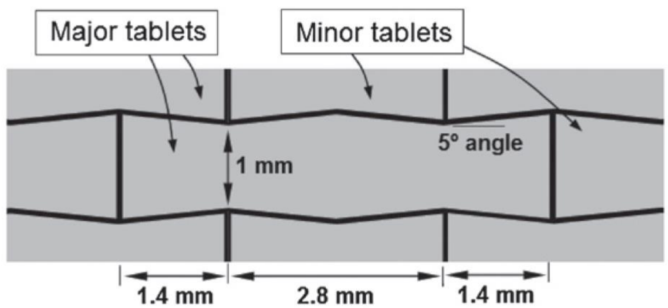



Figure 26. Schematic of the bioinspired tablets with 'dovetail' ends [129].

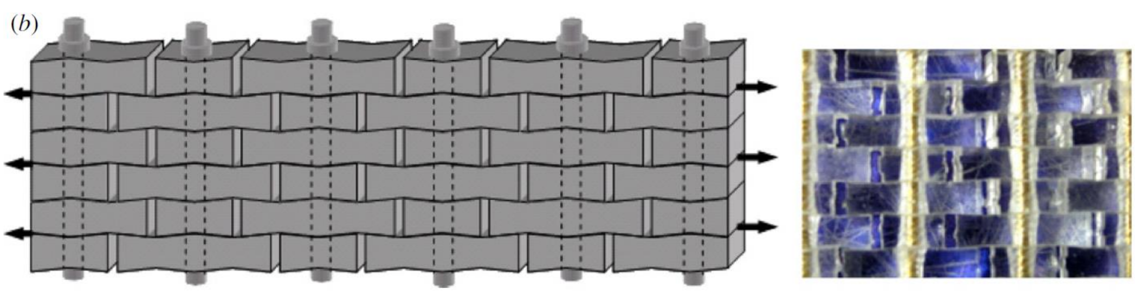



Figure 27. (Left) Schematic of the artificial nacre and tablet sliding in tension. (Right) The transparent PMMA platelets and the brass bolts, imitating the platelet sliding and mineral bridges of nacre [129].

### *Sacrificial Domains*

The sacrificial domain principle relates the usage of such domains to improving material toughness. Often times the sacrificial domains make use of unraveling or unfolding in order to dissipate energy. Nacre and bone have loosely packed modules of connective proteins that improve toughness through molecular friction during domain unfolding [26]. Others have observed mineralized collagen fibers making use of bioapatite as sacrificial bonds that contribute to bone's toughness [105]. Within the articles used for analysis, spider silks and muscle tissues were also found to utilize sacrificial domains to improve material toughness by unfolding the crystals that reinforce the macromolecules [99]. Alternatively, materials such as sponge spicules and armored fish scales have sacrificial outer layers that dissipate energy to protect the inner layers [92, 119].

The selected bioinspired material utilizes the concept of loosely packed modules, as found in titin and connective proteins in soft and hard tissues [19]. The modules unfold sequentially when deformed in order to maintain high strength and large elongation, which result in high material toughness. The modular structures are reversibly unfolding, which can also improve the material's stiffness. The bioinspired material utilized a poly(*n*-butyl acrylate) network and a biomimetic cyclic modular cross-linker 4-ureido-2-pyrimidone (UPy) (Figure 28). For comparison, a poly(ethylene glycol) (PEG) dimethacrylate was used as a cross-linker, which has a similar unfolded length to the unfolded UPy cross-linker. At 6% cross-linking, the UPy sample's strength is 700% greater than that of the control PEG sample (Figure 29), and the elongation at failure was 0.8 mm/mm as compared to the PEG sample's 0.19 mm/mm. The modulus of the UPy sample was also continuously improving. The modular cross-linker allowed for improving tensile strength and modulus as the cross-linker density increased, without a loss of extensibility. For the PEG sample, increasing the cross-linker levels corresponded to a decrease in maximum elongation due to increased rigidity and decreased elasticity, as is characteristic of thermoset elastomers. Like the loosely packed modules of biological materials, the usage of modular cross-linkers that unfold under tension improved the bioinspired material's toughness.

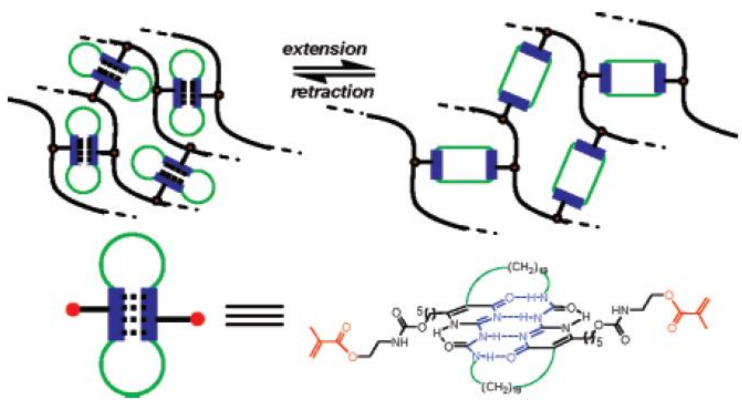



Figure 28. The bioinspired design concept containing an unfolding modular cross-linker of UPy [19].

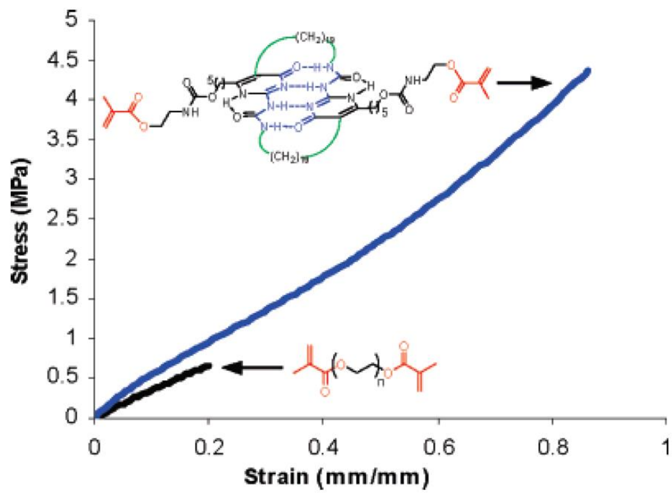



Figure 29. Stress-strain curves for the UPy sample and the control PEG sample at 6% cross-linking [19].

*Fiber Bundles*

The ninth principle employs fiber bundles. The fiber bundles can affect stiffness, density, strength, or ductility depending on the material. Fiber bundles are combined with the gradient principle in palm stems and bamboo. For palm stems, there are an increased number of larger diameter bundles near the periphery of the stem, which produces a radial gradient of both density and modulus [120]. Bamboo uses vascular bundles, which contain fibrous sclerenchyma cells and hollow vessels, in the bamboo culm. A greater number of these bundles are near the outer wall, which also results in a stiffness and density gradient [100, 114]. Bundles of fibers can also be found in yew [122]. The tail feathers of the toucan make use of fiber bundles that are oriented longitudinally, also making use of the fiber alignment principle [95]. When stressed, the bulk cortex fiber bundles are reorganized, which results in an increase in failure strains and ductility.

The examined bioinspired material used collagen bundles as inspiration in order to simulate carbon nanotube (CNT) bundles [130]. Carbon nanotubes have high strength and stiffness, but are difficult to use as a composite reinforcement due to their weak interfaces, misalignment, and inadequate distribution. To create CNT bundles, the authors looked to collagen, which has desired material properties. The collagen fibers have multiple length scales from collagen fibers to bundles of fibrils to bundles of tropocollagen molecules, which have an aligned quarterly-staggered pattern. The bioinspired CNT bundle structure mimicked the staggered packing feature with a controllable crosslink density. A schematic of the CNT bundles and their packing can be found in Figure 30. Using molecular mechanics simulations, the bioinspired bundles were found to have tensile strength that was an order of magnitude larger than the strongest conventional carbon fibers. The elastic modulus of the bioinspired CNT bundles was similar to the stiffest carbon fibers. Compared to conventional cross-linked CNT bundles, the bioinspired bundles had greater strength and ductility. Based on these results, the authors suggest that bioinspired CNT bundles can be used as fiber reinforcement for effective future nanocomposites.

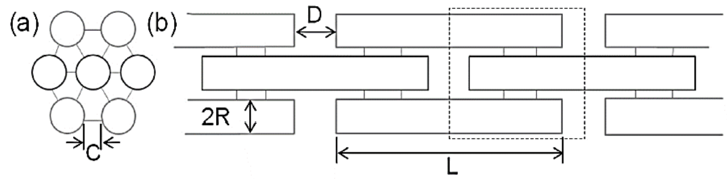



Figure 30. (Left) Cross-sectional schematic of the triangular lattice CNT bundle with the 7 CNTs. (Right) Front view of the half-staggered packing pattern composed of layers of CNTs [130].

### *Shell with Foam Core*

The final principle presented in this chapter utilizes a dense shell with a foam-like core. The structure is a specific form of the layered structure and mainly affects density, stiffness, and strength. A variety of biological materials employ this structure. One of the more well-known examples are feathers, which have a solid that envelopes a cellular core in order to reduce feather weight [5, 91]. Similarly, some bird beaks make use of the same structure. Toucan beaks have a closed-cell core with an outer beta-keratin shell in order to obtain adequate stiffness and low density [5]. Other natural materials with a solid outer shell and a foam-like core include porcupine quills, palm, bamboo, and bone [91, 120].

The selected bioinspired material took inspiration from the biological core-shell composites and specifically cited plant stems, bird beaks, and porcupine quills [131]. The bioinspired material’s application is orthopedic tissue engineering, which requires a biomaterial that has the required stiffness, strength, and necessary bioactivity in terms of permeability. Unlike the other bioinspired materials introduced throughout the chapter, this particular bioinspired material uses biological components and emphasizes bioactivity alongside the mechanical properties. For tissue engineering, there are two common material options, each with their own individual issues. One option is an electrospun polymer mat, which has the required stiffness and supports cell alignment, but has limited cell penetration. On the other hand, porous scaffold biomaterials have better permeability, but insufficient mechanical properties. The selected bioinspired material uses the core-shell principle in order to obtain the necessary stiffness, strength, and bioactivity. The material uses collagen-glycosaminoglycan (CG) for both the scaffold core and the membrane shell. The highly porous scaffold core has a lower relative density and is anisotropic, while the membrane shell has a higher relative density and tensile strength. Using an evaporative process, the membrane is first fabricated, and then the scaffold is integrated by freeze-drying and crosslinking, resulting in the composite seen in Figure 31. For a 155 μm thick membrane, the core-shell bioinspired composite's tensile moduli was 36 times greater than that of the CG scaffold. The density of the overall composite is less than that of the CG membrane. In terms of bioactivity, the bioinspired composite was found to have adequate permeability for tendon cell viability and proliferation. The bioinspired core-shell construct affected both stiffness and density, as anticipated based on the biological materials, but was also able to achieve the requirements necessary for bioactivity. Some of the identified design principles may be used for purposes beyond the expected structural materials.

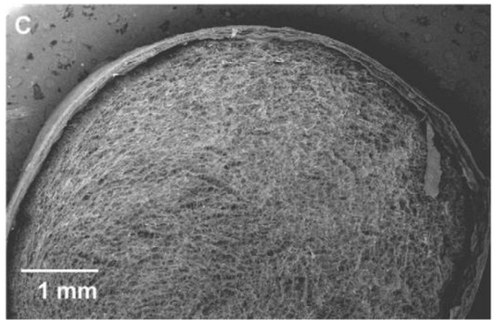



Figure 31. Cross-sectional SEM image of the scaffold-membrane CG biomaterial inspired by core-shell biological materials [131].

## Conclusion

In this chapter, the Stage 4 text mining tool was used to identify the ten principles mentioned most often within the four previously used sets of articles. The goal was to validate that the program identifies structural design principles and that the principles can be used to create bioinspired materials. For each of the selected principles, the 72 articles were examined to determine which biological materials utilize the principle and to further understand the structure-property relationships. Bioinspired materials designed using the principles were selected and assessed to determine how the structure affected the bioinspired materials' properties.

Many of the selected principles were in agreement with the "common design motifs" identified separately by Wegst et al. in their "Bioinspired structural materials" article. Of the top ten principles identified by the Stage 4 tool, seven of the structures were mentioned within their paper alongside many of the same expected property effects. For example, the article noted that "other examples of evolution-driven strategies are the use of highly porous architectures in materials that must combine light weight and stiffness, such as cancellous bone or bamboo" [2]. The strategy aligns with the text mining tool's top principle, which relates porous structures to stiffness and density. The tool was successful in identifying principles, as they were in agreement with those manually chosen by others doing research on biological and bioinspired materials.

Most of the principle structures were used as inspiration to develop new bioinspired materials. One of the principles did not yet have a corresponding bioinspired material and several of the others did not use the principle exactly. The differences are largely due to the difficulty in creating the bioinspired materials. Many of the biological materials use a variety of length scales from the nanoscale to the macroscale, and require certain orientations of components within a composite. Some of the processing and manufacturing required to complete these concepts are still in development. However, with further research and effort, the issues can be overcome. This is exemplified by the development of many materials inspired by nacre's brick-and-mortar structure. Since nacre is one of the most widely researched biological materials and much corresponding bioinspired materials research has also been completed, considerable advancements have been made in developing tough and strong materials based on the brick-and-mortar structure. Examining the bioinspired materials presented throughout the chapter, it is evident that there is potential and progress being made in successfully creating bioinspired materials from the other structural design principles as well.

One observation is that there are overlaps between principles. For example, porosity can be used to achieve a structural gradient effect, which then translates to property gradients. Another example is the core-shell principle. The architecture contains two layers, which is a form of the layered structure principle, and the core is foam-like, which can be considered a high porosity material. The usage of multiple principles may contribute to improved property effects. Other unique principles may also be used in conjunction in the development of bioinspired materials.

Overall, the Stage 4 text mining tool is successfully identifying structural design principles. A majority of the principles identified in this chapter have been used to create bioinspired materials, which affect the properties as predicted by the principle structure-property relationships. Based on the effective development of bioinspired materials using the more well-known principles, materials designers can now use the text mining tool to look towards the less common structural design principles. These principles may be less obvious to materials designers, but can still inspire new and creative bioinspired material developments.

# CHAPTER VIII

# CONCLUSIONS AND FUTURE WORK

Biological materials are able to achieve a wide range and combination of properties through the arrangement of the material's components. These biological materials are often more effective and better suited to their function than engineered materials, even with the limited set of components. By mimicking a biological material's component arrangement, or structure, man-made bioinspired materials can achieve improved properties as well. While considerable research has been conducted on biological materials, materials designers need to read through the research articles in order to identify the structural design principles, which can be time-intensive. The research in this dissertation centered on the creation of a text mining algorithm that allows materials designers to extract structural design principles much more quickly. Any materials text can be applied to the tool in order to retrieve short passages that describe biological material structures that affect a selected property or combination of properties.

The text mining problem is unique because the algorithm needs to identify passages describing structure-property relationships. While the properties are selected by the user and their associated terms are known, the biological material structures and the design principles are not known and cannot be used for the search. Properties are mentioned throughout materials articles, but often not in relationship to the material structures, which prevents the utilization of a simple property term search to identify the structural design principles. The difficulty with identifying the structure-property relationships is exacerbated by the variability in the English language and individual writing styles. A final added difficulty is a desire to create an algorithm with an underlying structure that is applicable to other domains as well.

The following sections give brief summaries of the approaches used and discuss the conclusions drawn from each stage of the research. Subsequently, the overall conclusions of the work are presented, along with a discussion of the successfully developed algorithms. Finally, the research contributions and recommendation for future work are reviewed.

## Conclusions: Stage 1 of Algorithm Development

The first iteration of the algorithm was an information retrieval tool developed to maximize the number of extracted structural design principles. Chapter 3 covered the algorithm development from the determination of the search term categories and the selection of the search terms themselves to the refinement of the program. The algorithm was based on searching for terms from a material property term set and the relationship term set, and required terms from each set to appear within two sentences of each other. The created tool achieved a very high recall (percentage of identified principles) of 98% for a new set of journal articles. The tool's identified principles were also found to begin repeating between different journal articles and biological materials, which may help materials designers by displaying different ways the principles can be applied. However, a shortcoming of the algorithm was that the percentage of relevant passages, or precision, was only 32%. While the Stage 1 algorithm would help a materials designer identify structural design principles more quickly than if they were to read through all the articles individually, the algorithm could be improved to further reduce the user's workload.

## Conclusions: Stage 2 of Algorithm Development

For the rest of algorithm development, the goal was to reduce the number of irrelevant passages returned by the program so that the tool's user would receive a greater percentage of useful information. To accomplish this goal, the text mining task of text classification was considered. Text classification automatically assigns documents to their respective classes. For this research, the documents were passages, and the classes were the "relevant passages" class and the "irrelevant passages" class. If passages have been sorted into their respective classes, the "irrelevant passages" class could be removed, thus improving the tool's precision. The first investigated classic text classification technique was the use of supervised machine learning classifiers, including both a Naïve Bayes and a maximum entropy classifier. Next, an assessment of the usage of statistical features to classify the passages was conducted in terms of the number of words per property sentence and per passage. The third technique analyzed the parts of speech employed within the relevant and irrelevant passages. None of the three techniques were consistently successful in classifying the passages. The research problem at hand is different enough from common text classification applications that the classic techniques and well-researched classification algorithms cannot be used in the design principle algorithm. One of the issues with classifying the passages is that there are different principles, materials, and authors introduced within the relevant and irrelevant classes, making it difficult for the program to find unifying themes within each class. Another issue is that multiple relevant and irrelevant passages may be identified within individual articles. Since the passages share the same biological materials and authors, who each generally use a consistent writing style throughout their paper, preset techniques have difficulty identifying differences between the two classes. This contrasts with text classification work that uses full documents, articles, or abstracts for classification, which reduces the amount of overlap between the classes.

## Conclusions: Stage 3 of Algorithm Development

In the third stage of algorithm development, manual rule-based classification was employed. The technique is a less commonly used form of text classification, and requires the manual identification of rules and patterns based on combinations of words or other features. A technically skilled domain expert with good pattern identification skills is needed in order to create an accurate algorithm [59]. Chapter 5 described the approach taken to identifying the patterns and then discussed the patterns found within the irrelevant and relevant passages. The patterns led to updated property search terms and the addition of exclusion terms, stand-alone terms, and identifier terms to the Stage 3 algorithm. The research conducted during Stage 3 also found that three-sentence passages were suitable to identify a majority of the principles without returning an excess amount of unnecessary text. The Stage 1 algorithm was revised to take into account the change in passage length for a more consistent comparison between the two algorithms. Both the Stage 3 and the revised Stage 1 algorithms resulted in improved precision, but reduced recall over the original Stage 1 algorithm. The reduced recall was expected because the two metrics are inversely related. The Stage 3 tool achieved a precision of 45% and a recall of 80%. The tool was slightly outperformed in terms of the two metrics by the Stage 1 tool with its 46% precision and 82% recall. The revised Stage 1 tool's results were on par with other initial text classification algorithm work. The Stage 3 tool returned 22% fewer passages, which reduces the amount of reading required by a materials designer. Compared to the original Stage 1 tool, the revised Stage 1 tool saw an improvement in terms of precision, even with the reduced passage length. The ratio of useful passages to total passages for the Stage 1 algorithms increased from approximately 1:3 to nearly 1:2. The revised Stage 1 tool would be considered superior to the Stage 3 tool if a materials designer desires more sources of inspiration and information. The Stage 3 tool may be preferred if the tool's user wants to skim through potential principles more quickly.

## Conclusions: Stage 4 of Algorithm Development

The final stage of algorithm development was completed in an attempt to further reduce the effort required from materials designers to identify structural design principles. While Stage 3 aimed to improve precision, retaining a high recall was still important during algorithm development. Since versions of the tool that achieve high recall have been developed, Stage 4 concentrated on creating a final version of the tool to further improve precision with less concern for recall. With the goal of reducing the end user's workload, the emphasis is shifted towards returning more useful passages over identifying all possible principles. During the fourth stage, a second round of manual rule-based classification took place, which resulted in the identification of additional patterns in the irrelevant and relevant classes. The patterns were stricter than those in Stage 3 so that their application excludes a greater number of relevant passages, but also filter more of the irrelevant passages. From the irrelevant passages, six patterns emerged. The irrelevant passage patterns included the usage of numerical figures, ratios, and the word "not" prior to a relationship term. The other patterns related to the irrelevant passage topics, which were detected using sets of method, calculation, and result terms. The relevant passages were identified by six different templates that made use of the modification, causation, correlation, and comparison term sets. A second part-of-speech analysis took place with a close examination of the relevant passage term sets, although the analysis did not yield patterns for the Stage 4 algorithm. Additionally, a passage activity was completed by graduate-level mechanical engineers to leverage their pattern identification abilities. The patterns they found were already in use or considered for the final algorithm. However, their efforts did affirm the research problem's level of difficulty and the need for materials domain expertise.

Based on a comparison to the revised Stage 1 algorithm, the Stage 4 algorithm's overall ability and accuracy was improved. The Stage 4 algorithm successfully improved the precision to 63%, although the recall was reduced to 65% due to the strict filtering rules. In additional to the precision improvement, the Stage 4 algorithm resulted in a reduction of approximately 74% of the passages and the word count, making the Stage 4 results significantly faster to peruse and reducing the work load required from the tool's user. The Stage 4 algorithm's manually identified patterns and templates were successful in eliminating many procedural descriptions and basic results listings, leading to a more readable passage set. However, it was noted that if the Stage 4 tool returned too few passages for a particular property, it may be necessary to run the corpus through the revised Stage 1 algorithm in order to expand the results for additional information and structural design principles. Once again, the two algorithms have different uses, but both resulting sets of passages would be faster for a materials designer to read through than full articles when searching for potential biological materials design principles.

## Overall Conclusions

The research successfully developed algorithms that return passages with structural design principles and satisfy the goal of reducing materials designer effort in identifying the principles. Each stage of the research worked to improve upon the previous work to create a tool that is more useful to materials designers. Two final text mining tools were deemed most useful: either the revised Stage 1 tool or the Stage 4 tool can be selected based on the materials designer's needs. The results from both algorithms were good compared to other text classification studies that require domain expertise. As mentioned in the background chapter, McKnight and Srinavasan's best classifier approach achieved a recall of 49% and a precision of 82% for automatically classifying sentences by type [67]. Like the work presented in this chapter, their work was the first in their field. Their results skewed towards precision, while this dissertation work was towards recall, but the values are similar. The identification of passages describing structure-property relationships from biological materials articles may be more difficult than classifying biomedical sentences into introduction, method, result, and conclusion types because this dissertation work requires the extraction of passages and the identification of new principles and structures. Also, the materials domain does not benefit from having a language system equivalent to the biomedical field's UMLS.

If the designer is willing to put in a more effort in order to obtain an abundance of sources of inspiration and maximize the number of principles returned, then they should consider the revised Stage 1 tool. In Stage 1, the algorithm is more open with only biological, property, and relationship term search requirements, allowing for the retrieval of many passages. If a designer wanted to quickly survey the principles with an emphasis on reducing the amount of reading and effort, then the Stage 4 tool is recommended. Stage 4's algorithm is far more restrictive with the addition of the relevant passage templates requiring combinations of modification, correlation, comparison, and causation terms with the property and stand-alone search terms. The Stage 4 algorithm also filtered many of the passages with the use of exclusion terms, method terms, calculation terms, result terms, numerical figures, and forward slashes. The Stage 4 algorithm removed many of the irrelevant passages, although various relevant passages and potential design principles were removed as well. The strict Stage

4 algorithm resulted in a greatly reduced number of retrieved passages in combination with an improvement in the percentage of relevant passages. Through the validation chapter, it was shown that the text mining tools are successfully identifying structural design principles, and that the principles can be used for the creation of new bioinspired materials with improved properties.

## Research Contributions

On a more abstract level, this research's contribution is the development of an algorithm that can identify relationships within text. The algorithm can be used to return passages describing causal relationships, correlations, and comparisons so that the algorithm user can quickly locate important information in the text. The manually identified rules used in the final algorithm were carefully crafted to be specific enough for relevant design principle classification, but not so specific as to result in a brittle algorithm. While some of the search term sets are particular to the specific application, the fundamental algorithm structure and patterns are more generalizable. With the removal of the biological search terms, the tool can be expanded to identify general materials structure-property relationships. By modifying the property-specific terms, the algorithm may be used for topics outside of the materials domain.

The technical contribution of this work is the creation of a tool that utilizes text mining in order to assist with bioinspired materials design. Through the refined set of parameters and algorithms, the tool extracts passages describing structural design principles from materials text for a specified property or combination of properties. By identifying structural design principles, the text mining tool can aid materials designers in developing new and innovative materials with improved properties. A distinct advantage of the tool is that it is not simply a database, but can be built on and expanded in the future. The materials corpus can easily be updated with additional articles and text without any change to the algorithm itself, allowing the tool to identify additional principles as research into biological materials progresses. Methods to identify the property-specific terms required for the program have been established for the property search terms, stand-alone terms, and exclusion terms. One can easily apply the methods to other material properties to advance the program.

## Recommendations for Future Work

The completed research has initiated the utilization of text mining to help materials designers identify sources of inspiration for bioinspired materials. There are several avenues of further work that can be completed to further assist the designers. In this section, recommendations for future work in the area of text mining tool advancement and a design principle database are discussed.

### *Tool Advancement: Additional Properties*

To bring the tool beyond structural material applications, additional material properties can be integrated into the tool. Currently, the tool allows for the property selection of density, ductility, hardness, stiffness, strength, and toughness. Properties can be added by simply following the methods used to identify the property-specific terms required for the program. The property search terms were initially selected in Stage 1, and a refining of the terms was discussed in Stage 3. The process used for the identification of stand-alone terms and exclusion terms were both described in Stage 3. To take advantage of biological materials' unique abilities, various properties can be added, including adhesion, sensing, self-repair, wear resistance, self-cleaning, reactivity, and thermal modulation. Ideally, the program will be enhanced to provide a materials designer with an easy means of identifying biological structure-property design principles for any notable property. Beyond properties, the tool may be used to look at other biological material abilities as well such as manufacturing, actuation, and bioactivity for biomedical implants and applications.

### *Tool Advancement: Passage Categorization*

Additional text mining tasks and tools can be used to automatically organize the results for the tool's users. The algorithm can be updated to group the returned passages by structural design principle by utilizing a clustering algorithm, which identifies similar objects in a data set [132]. Since clustering techniques often make use of term frequency, the text would first need to be pre-processed in order to remove commonly used general terms such as "the", "and", and "to". Terms common in general materials papers may also want to be removed in order to focus the results on areas unique to biological materials. The search terms used by the general text mining tool should also be eliminated since those terms are expected to appear often.

Based on initial investigation into passage categorization, a variety of issues require consideration. For general text clustering, the main issues are the high volume of data, the high dimensionality due to the large number of unique terms, and the unknown number of clusters required prior to categorization [133]. General partition-based, hierarchical, and density-based clustering, which are standard clustering techniques, cannot be used due to these issues. Additional issues arise due to the nature of the structural design principle text mining tool. The algorithm returns multi-sentence passages that may have more than one principle per passage. Another issue is that structural design principles may be worded a variety of ways (e.g. porous, pores, and porosity), which prevents the use of a basic co-occurrence matrix.

Frequent itemset-based clustering algorithms can alleviate many of these issues. These algorithms are commonly used for categorizing news stories or clustering query results. The thought is that passages in the same cluster will share a set of terms, or itemset, that occur more frequently in that cluster than in other clusters [133]. There are not specific terms that are required in order for a passage to belong to a cluster. By using itemsets rather than individual terms for clustering, the algorithm should be able to manage the various principle descriptors. Further study into frequent itemset-based clustering algorithms is required in order to select an algorithm and determine how to best implement it into the current text mining algorithm.

*Tool Advancement: Multi-Document Summarization*

Following passage categorization, the tool can be further advanced by adding the text mining task of multi-document summarization. Algorithms have been created that automatically summarize multiple documents concerning the same topic. By utilizing such algorithms, the tool may be able to summarize each of the cluster principles for the user. The desired end product would be the output of structural design principle key terms and a concise summary or description of the structure and how it affects the properties.

*Design Principle Database*

A final avenue of future work would be the development of a structural design principle database. While one of the advantages of the text mining tool is its ability to analyze different text in order to identify new principles, some materials designers may want a quick resource of predetermined design principles that they can reference. A database can be created that contains the structural design principles by property. For each principle, the database can provide detailed information on the structure-property relationship, biological materials that use the principles, references to articles that describe the principles, and bioinspired materials that have been created based on the principles.

# REFERENCES

[1] Fabritius, H.-O., Sachs, C., Triguero, P. R., and Raabe, D., 2009, "Influence of structural principles on the mechanics of a biological fiber-based composite material with hierarchical organization: the exoskeleton of the lobster homarus americanus," Advanced Materials, 21(4), pp. 391-400.

[2] Wegst, U. G., Bai, H., Saiz, E., Tomsia, A. P., and Ritchie, R. O., 2015, "Bioinspired structural materials," Nature Materials, 14(1), pp. 23-36.

[3] Kantardzic, M., 2011, *Data Mining: Concepts, Models, Methods, and Algorithms*, John Wiley & Sons, Inc., Hoboken, NJ.

[4] Lee, N., Horstemeyer, M., Rhee, H., Nabors, B., Liao, J., and Williams, L. N., 2014, "Hierarchical multiscale structure–property relationships of the red-bellied woodpecker (Melanerpes carolinus) beak," Journal of the Royal Society Interface, 11(96), p. 20140274.

[5] Meyers, M. A., Chen, P.-Y., Lopez, M. I., Seki, Y., and Lin, A. Y., 2011, "Biological materials: a materials science approach," Journal of the Mechanical Behavior of Biomedical Materials, 4(5), pp. 626-657.

[6] Vincent, J. F., 1990, *Structural Biomaterials: Revised Edition*, Princeton University Press, Princeton, NJ.

[7] Ortiz, C., Eriksson, O., and Klintenberg, M., 2009, "Data mining and accelerated electronic structure theory as a tool in the search for new functional materials," Computational Materials Science, 44(4), pp. 1042-1049.

[8] Fratzl, P., 2007, "Biomimetic materials research: what can we really learn from nature's structural materials?," Journal of the Royal Society Interface, 4(15), pp. 637-642.

[9] Wegst, U. G., and Ashby, M. F., 2004, "The mechanical efficiency of natural materials," Philosophical Magazine, 84(21), pp. 2167-2186.

[10] Aizenberg, J., Weaver, J. C., Thanawala, M. S., Sundar, V. C., Morse, D. E., and Fratzl, P., 2005, "Skeleton of Euplectella sp.: structural hierarchy from the nanoscale to the macroscale," Science, 309(5732), pp. 275-278.

[11] Bechtle, S., Ang, S. F., and Schneider, G. A., 2010, "On the mechanical properties of hierarchically structured biological materials," Biomaterials, 31(25), pp. 6378-6385.

[12] Sanchez, C., Arribart, H., and Guille, M. M. G., 2005, "Biomimetism and bioinspiration as tools for the design of innovative materials and systems," Nature Materials, 4(4), pp. 277-288.

[13] Vincent, J. F., 2008, "Biomimetic materials," Journal of Materials Research, 23(12), pp. 3140-3147.

[14] Gao, H., Ji, B., Jäger, I. L., Arzt, E., and Fratzl, P., 2003, "Materials become insensitive to flaws at nanoscale: lessons from nature," Proceedings of the National Academy of Sciences, 100(10), pp. 5597-5600.

[15] Dume, B., 2006, "Ceramics come out of the cold," Physics World, http://physicsworld.com/cws/article/news/2006/jan/26/ceramics-come-out-of-the-cold.

[16] Bonderer, L. J., Studart, A. R., and Gauckler, L. J., 2008, "Bioinspired design and assembly of platelet reinforced polymer films," Science, 319(5866), pp. 1069-1073.

[17] Geim, A., Grigorieva, S. V. D. I. V., Novoselov, K., Zhukov, A., and Shapoval, S. Y., 2003, "Microfabricated adhesive mimicking gecko foot-hair," Nature Materials, 2(7), pp. 461-463.

[18] Chaplin, C. R., Gordon, J. E., and Jeronimidis, G., 1983, "Composite material," US Patent 4409274 A, 11 Oct 1983.

[19] Kushner, A. M., Gabuchian, V., Johnson, E. G., and Guan, Z., 2007, "Biomimetic design of reversibly unfolding cross-linker to enhance mechanical properties of 3D network polymers," Journal of the American Chemical Society, 129(46), pp. 14110-14111.

[20] 2015, "Nacre," Wikipedia, https://en.wikipedia.org/wiki/Nacre.

[21] 2016, "Web of Science Citation Report," Thomson Reuters, http://webofscience.com.

[22] Gibson, L. J., 2012, "The hierarchical structure and mechanics of plant materials," Journal of the Royal Society Interface, 9(76), pp. 2749-2766.

[23] Gibson, L. J., and Ashby, M. F., 1999, *Cellular Solids: Structure and Properties*, Cambridge University Press, New York, NY.

[24] Miserez, A., Weaver, J. C., Pedersen, P. B., Schneeberk, T., Hanlon, R. T., Kisailus, D., and Birkedal, H., 2009, "Microstructural and biochemical characterization of the nanoporous sucker rings from Dosidicus gigas," Advanced Materials, 21(4), pp. 401-406.

[25] Wegst, U. G., and Ashby, M. F., 2007, "The structural efficiency of orthotropic stalks, stems and tubes," Journal of Materials Science, 42(21), pp. 9005-9014.

[26] Harrington, M. J., and Waite, J. H., 2009, "How nature modulates a fiber's mechanical properties: mechanically distinct fibers drawn from natural mesogenic block copolymer variants," Advanced Materials, 21(4), pp. 440-444.

[27] Vincent, J. F., Bogatyreva, O. A., Bogatyrev, N. R., Bowyer, A., and Pahl, A.-K., 2006, "Biomimetics: its practice and theory," Journal of the Royal Society Interface, 3(9), pp. 471-482.

[28] Morgan, D., and Ceder, G., 2005, "Data mining in materials development," *Handbook of Materials Modeling*, Springer, Dordrecht, The Netherlands, pp. 395-421.

[29] Chen, P.-Y., McKittrick, J., and Meyers, M. A., 2012, "Biological materials: functional adaptations and bioinspired designs," Progress in Materials Science, 57(8), pp. 1492-1704.

[30] McKittrick, J., Chen, P.-Y., Tombolato, L., Novitskaya, E., Trim, M., Hirata, G., Olevsky, E., Horstemeyer, M., and Meyers, M., 2010, "Energy absorbent natural materials and bioinspired design strategies: a review," Materials Science and Engineering: C, 30(3), pp. 331-342.

[31] Vincent, J. F. V., and Mann, D. L., 2002, "Systematic technology transfer from biology to engineering," Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 360(1791), pp. 159-173.

[32] Tsenn, J., McAdams, D. A., and Linsey, J. S., 2015, "A comparison of mechanical engineering and biology students' ideation and bioinspired design abilities," *Design Computing and Cognition '14*, Springer, Switzerland, pp. 645-662.

[33] Glier, M. W., Tsenn, J., Linsey, J. S., and McAdams, D. A., 2014, "Evaluating the directed intuitive approach for bioinspired design," Journal of Mechanical Design, 136(7), p. 071012.

[34] Chakrabarti, A., Sarkar, P., Leelavathamma, B., and 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), pp. 113-132.

[35] Goel, A. K., McAdams, D. A., and Stone, R. B., 2013, *Biologically Inspired Design: Computational Methods and Tools*, Springer, London.

[36] Nagel, J. K., Stone, R. B., and McAdams, D. A., 2010, "An engineering-to-biology thesaurus for engineering design," *ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Montreal, Quebec, Canada.

[37] 2013, "AskNature," The Biomimicry 3.8 Institute, www.asknature.org.

[38] Vattam, S., Wiltgen, B., Helms, M., Goel, A. K., and Yen, J., 2011, "DANE: fostering creativity in and through biologically inspired design," *Design Creativity 2010*, Springer, London, pp. 115-122.

[39] Chiu, I., and Shu, L., 2007, "Biomimetic design through natural language analysis to facilitate cross-domain information retrieval," Artificial Intelligence for Engineering Design, Analysis, and Manufacturing, 21(1), pp. 45-59.

[40] Chiu, I., and Shu, L. H., 2004, "Natural language analysis for biomimetic design," *ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Salt Lake City, UT.

[41] Hirtz, J., Stone, R. B., McAdams, D. A., Szykman, S., and Wood, K. L., 2002, "A functional basis for engineering design: reconciling and evolving previous efforts," Research in Engineering Design, 13(2), pp. 65-82.

[42] Kaiser, M. K., Farzaneh, H. H., and Lindemann, U., 2013, "BIOscrabble - extraction of biological analogies out of large text sources," *International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, Vilamoura, Portugal.

[43] Deldin, J.-M., and Schuknecht, M., 2014, "The AskNature database: enabling solutions in biomimetic design," *Biologically Inspired Design*, Springer, London, pp. 17-27.

[44] Vandevenne, D., Verhaegen, P.-A., Dewulf, S., and Duflou, J. R., 2010, "A scalable approach for the integration of large knowledge repositories in the biologically-inspired design process," *International Conference on Engineering Design*, Copenhagen, Denmark.

[45] Vandevenne, D., Verhaegen, P.-A., Dewulf, S., and Duflou, J. R., 2015, "A scalable approach for ideation in biologically inspired design," Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 29(01), pp. 19-31.

[46] Vandevenne, D., Verhaegen, P.-A., Dewulf, S., and Duflou, J. R., 2016, "SEABIRD: Scalable search for systematic biologically inspired design," Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 30(01), pp. 78-95.

[47] Altshuller, G. S., Shulyak, L., and Rodman, S., 1999, *The Innovation Algorithm: TRIZ, Systematic Innovation and Technical Creativity*, Technical Innovation Center, Inc., Worcester, MA.

[48] 2016, "Part II Mining Functions," Oracle, http://docs.oracle.com/cd/B28359_01/datamine.111/b28129/text.htm.

[49] Cohen, A. M., and Hersh, W. R., 2005, "A survey of current work in biomedical text mining," Briefings in Bioinformatics, 6(1), pp. 57-71.

[50] Niezgoda, S. R., Kanjarla, A. K., and Kalidindi, S. R., 2013, "Novel microstructure quantification framework for databasing, visualization, and analysis of microstructure data," Integrating Materials and Manufacturing Innovation, 2(1), pp. 1-27.

[51] Chalk, A. J., Beck, B., and Clark, T., 2001, "A quantum mechanical/neural net model for boiling points with error estimation," Journal of Chemical Information and Computer Sciences, 41(2), pp. 457-462.

[52] Tuchbreiter, A., and Muelhaupt, R., 2001, "The polyolefin challenges: catalyst and process design, tailor-made materials, high-throughput development and data mining," Macromolecular Symposia, 173(1), pp. 1-20.

[53] Rajan, K., 2008, "Combinatorial materials sciences: Experimental strategies for accelerated knowledge discovery," Annual Review of Materials Research, 38(2008), pp. 299-322.

[54] Curtarolo, S., Morgan, D., Persson, K., Rodgers, J., and Ceder, G., 2003, "Predicting crystal structures with data mining of quantum calculations," Physical Review Letters, 91(13), p. 135503.

[55] Fischer, C. C., Tibbetts, K. J., Morgan, D., and Ceder, G., 2006, "Predicting crystal structure by merging data mining with quantum mechanics," Nature Materials, 5(8), pp. 641-646.

[56] Stephenson, L. T., Moody, M. P., Liddicoat, P. V., Ringer, S. P., Byers, S., Rafferty, A., Cerezo, A., Davin, L., Clark, P., and Evans, F., 2007, "New techniques for the analysis of fine-scaled clustering phenomena within atom probe tomography (APT) data," Microscopy and Microanalysis, 13(6), pp. 448-463.

[57] Ashby, M. F., 2005, *Materials Selection in Mechanical Design, Third Edition*, Butterworth-Heinemann, Burlington, MA.

[58] Lin, J., Quan, D., Sinha, V., Bakshi, K., Huynh, D., Katz, B., and Karger, D. R., 2003, "What makes a good answer? The role of context in question answering," *International Conference on Human-Computer Interaction*, Zurich, Switzerland.

[59] Manning, C. D., Raghavan, P., and Schütze, H., 2008, *Introduction to Information Retrieval*, Cambridge University Press Cambridge, Cambridge, England.

[60] Jurafsky, D., and Manning, C., 2012, "Text Classification," Stanford Natural Language Processing, https://www.coursera.org/course/nlp.

[61] Donaldson, I., Martin, J., De Bruijn, B., Wolting, C., Lay, V., Tuekam, B., Zhang, S., Baskin, B., Bader, G. D., and Michalickova, K., 2003, "PreBIND and Textomy–mining the biomedical literature for protein-protein interactions using a support vector machine," BMC Bioinformatics, 4(11), p. 1-13.

[62] Liu, F., Jenssen, T.-K., Nygaard, V., Sack, J., and Hovig, E., 2004, "FigSearch: a figure legend indexing and classification system," Bioinformatics, 20(16), pp. 2880-2882.

[63] Raychaudhuri, S., Schütze, H., and Altman, R. B., 2002, "Using text analysis to identify functionally coherent gene groups," Genome Research, 12(10), pp. 1582-1590.

[64] Wren, J. D., and Garner, H. R., 2004, "Shared relationship analysis: ranking set cohesion and commonalities within a literature-derived relationship network," Bioinformatics, 20(2), pp. 191-198.

[65] 2015, "Unified Medical Language System," NIH U.S. National Library of Medicine, https://www.nlm.nih.gov/research/umls/.

[66] Dobrokhotov, P. B., Goutte, C., Veuthey, A.-L., and Gaussier, E., 2003, "Combining NLP and probabilistic categorisation for document and term selection for Swiss-Prot medical annotation," Bioinformatics, 19(suppl 1), pp. i91-i94.

[67] McKnight, L., and Srinivasan, P., 2003, "Categorization of sentence types in medical abstracts," *American Medical Informatics Association Symposium*, Washington, D.C.

[68] 2013, "Thesaurus.com," Dictionary.com, http://www.thesaurus.com/.

[69] 1998-2015, "Random Sequence Generator," https://www.random.org/sequences.

[70] Princeton University, 2010, "About WordNet"," WordNet, http://wordnet.princeton.edu.

[71] Hölscher, C., and Strube, G., 2000, "Web search behavior of Internet experts and newbies," Computer Networks, 33(1), pp. 337-346.

[72] Callister, W. D. J., 2006, *Materials Science and Engineering: An Introduction, Seventh Edition*, John Wiley & Sons, Inc., New York, NY.

[73] 2009, Advanced Materials: Biological and Biomimetic Materials, 21(4), pp. 379-494.

[74] Porter, D., and Vollrath, F., 2009, "Silk as a biomimetic ideal for structural polymers," Advanced Materials, 21(4), pp. 487-492.

[75] Keten, S., Xu, Z., and Buehler, M. J., 2011, "Triangular core as a universal strategy for stiff nanostructures in biology and biologically inspired materials," Materials Science and Engineering: C, 31(4), pp. 775-780.

[76] Manning, C., 2015, "Maxent Models and Discriminative Estimation," CS 124: From Languages to Information, https://web.stanford.edu/class/cs124/lec/Maximum_Entropy_Classifiers.pdf.

[77] Zhang, T., and Oles, F. J., 2001, "Text categorization based on regularized linear classification methods," Information Retrieval, 4(1), pp. 5-31.

[78] Sachan, L., 2015, "Logistic Regression vs Decision Trees vs SVM: Part II," Edvancer Eduventures, http://www.edvancer.in/.

[79] 2011, Journal of the Mechanical Behavior of Biomedical Materials, 4(5), pp. 625-826.

[80] 2011, Materials Science and Engineering: C, 31(4), pp. 713-820.

[81] Manning, C., and Klein, D., 2003, "Optimization, maxent models, and conditional estimation without magic," *Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, Edmonton, Canada.

[82] Rafferty, A., Kleeman, A., Finkel, J., and Manning, C., 2003, Stanford Classifer [Software], available from http://nlp.stanford.edu/software/classifier.html.

[83] Forman, G., 2007, "Feature selection for text classification," *Computational Methods of Feature Selection*, Taylor & Francis Group, LLC, Boca Raton, FL.

[84] Glier, M., 2013, "Machine Learning Based Classification of Textual Stimuli to Promote Ideation in Bioinspired Design," Ph.D. thesis, Texas A&M University.

[85] Lee, S., Novitskaya, E. E., Reynante, B., Vasquez, J., Urbaniak, R., Takahashi, T., Woolley, E., Tombolato, L., Chen, P.-Y., and McKittrick, J., 2011, "Impact testing of structural biological materials," Materials Science and Engineering: C, 31(4), pp. 730-739.

[86] Toutanova, K., Klein, D., Manning, C. D., and Singer, Y., 2003, "Feature-rich part-of-speech tagging with a cyclic dependency network," *Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, Edmonton, Canada.

[87] Marcus, M. P., Marcinkiewicz, M. A., and Santorini, B., 1993, "Building a large annotated corpus of English: The Penn Treebank," Computational Linguistics, 19(2), pp. 313-330.

[88] Yeh, A. S., Hirschman, L., and Morgan, A. A., 2003, "Evaluation of text data mining for database curation: lessons learned from the KDD Challenge Cup," Bioinformatics, 19(suppl 1), pp. i331-i339.

[89] Miserez, A., Weaver, J. C., Pedersen, P. B., Schneeberk, T., Hanlon, R. T., Kisailus, D., and Birkedal, H., 2009, "Microstructural and biochemical characterization of the nanoporous sucker rings from Dosidicus gigas," Advanced Materials, 21(4), pp. 401-406.

[90] Tsenn, J., Linsey, J. S., and McAdams, D. A., 2014, "Development of a search tool to identify structural design principles for bioinspired materials design," *ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Buffalo, NY.

[91] Weiss, I. M., and Kirchner, H. O., 2011, "Plasticity of two structural proteins: alpha-collagen and beta-keratin," Journal of the Mechanical Behavior of Biomedical Materials, 4(5), pp. 733-743.

[92] Song, J., Ortiz, C., and Boyce, M. C., 2011, "Threat-protection mechanics of an armored fish," Journal of the Mechanical Behavior of Biomedical Materials, 4(5), pp. 699-712.

[93] Abdel-Wahab, A. A., Alam, K., and Silberschmidt, V. V., 2011, "Analysis of anisotropic viscoelastoplastic properties of cortical bone tissues," Journal of the Mechanical Behavior of Biomedical Materials, 4(5), pp. 807-820.

[94] de Souza, G. B., de Lima, G. G., Kuromoto, N. K., Soares, P., Lepienski, C. M., Foerster, C. E., and Mikowski, A., 2011, "Tribo-mechanical characterization of rough, porous and bioactive Ti anodic layers," Journal of the Mechanical Behavior of Biomedical Materials, 4(5), pp. 796-806.

[95] Bodde, S., Meyers, M., and McKittrick, J., 2011, "Correlation of the mechanical and structural properties of cortical rachis keratin of rectrices of the Toco Toucan (Ramphastos toco)," Journal of the Mechanical Behavior of Biomedical Materials, 4(5), pp. 723-732.

[96] Barber, A. H., Lu, D., and Pugno, N. M., 2015, "Extreme strength observed in limpet teeth," Journal of the Royal Society Interface, 12(105), p. 20141326.

[97] Checa, A. G., Bonarski, J. T., Willinger, M. G., Faryna, M., Berent, K., Kania, B., González-Segura, A., Pina, C. M., Pospiech, J., and Morawiec, A., 2013, "Crystallographic orientation inhomogeneity and crystal splitting in biogenic calcite," Journal of the Royal Society Interface, 10(86), p. 20130425.

[98] Chen, F., Porter, D., and Vollrath, F., 2012, "Structure and physical properties of silkworm cocoons," Journal of the Royal Society Interface, 9(74), pp. 2299-2308.

[99] De Tommasi, D., Millardi, N., Puglisi, G., and Saccomandi, G., 2013, "An energetic model for macromolecules unfolding in stretching experiments," Journal of the Royal Society Interface, 10(88), p. 20130651.

[100] Dixon, P. G., and Gibson, L. J., 2014, "The structure and mechanics of Moso bamboo material," Journal of the Royal Society Interface, 11(99), p. 20140321.

[101] Erko, M., Younes-Metzler, O., Rack, A., Zaslansky, P., Young, S. L., Milliron, G., Chyasnavichyus, M., Barth, F. G., Fratzl, P., and Tsukruk, V., 2015, "Micro-and nano-structural details of a spider's filter for substrate vibrations: relevance for low-frequency signal transmission," Journal of the Royal Society Interface, 12(104), p. 20141111.

[102] Grawe, I., Wolff, J. O., and Gorb, S. N., 2014, "Composition and substrate-dependent strength of the silken attachment discs in spiders," Journal of the Royal Society Interface, 11(98), p. 20140477.

[103] Klein, M.-C. G., and Gorb, S. N., 2012, "Epidermis architecture and material properties of the skin of four snake species," Journal of the Royal Society Interface, 9(76), pp. 3140-3155.

[104] Laurent, C. M., Palmer, C., Boardman, R. P., Dyke, G., and Cook, R. B., 2014, "Nanomechanical properties of bird feather rachises: exploring naturally occurring fibre reinforced laminar composites," Journal of the Royal Society Interface, 11(101), p. 20140961.

[105] Liu, Y., Thomopoulos, S., Chen, C., Birman, V., Buehler, M. J., and Genin, G. M., 2014, "Modelling the mechanics of partially mineralized collagen fibrils, fibres and tissue," Journal of the Royal Society Interface, 11(92), p. 20130835.

[106] Lucas, P. W., Gaskins, J. T., Lowrey, T. K., Harrison, M. E., Morrogh-Bernard, H. C., Cheyne, S. M., and Begley, M. R., 2011, "Evolutionary optimization of material properties of a tropical seed," Journal of the Royal Society Interface, p. 20110188.

[107] Meyer, A., Pugno, N. M., and Cranford, S. W., 2014, "Compliant threads maximize spider silk connection strength and toughness," Journal of the Royal Society Interface, 11(98), p. 20140561.

[108] Pasquini, L., Molinari, A., Fantazzini, P., Dauphen, Y., Cuif, J.-P., Levy, O., Dubinsky, Z., Caroselli, E., Prada, F., and Goffredo, S., 2015, "Isotropic microscale mechanical properties of coral skeletons," Journal of the Royal Society Interface, 12(106), p. 20150168.

[109] Rafsanjani, A., Stiefel, M., Jefimovs, K., Mokso, R., Derome, D., and Carmeliet, J., 2014, "Hygroscopic swelling and shrinkage of latewood cell wall micropillars reveal ultrastructural anisotropy," Journal of the Royal Society Interface, 11(95), p. 20140126.

[110] Rajkhowa, R., Kaur, J., Wang, X., and Batchelor, W., 2015, "Intrinsic tensile properties of cocoon silk fibres can be estimated by removing flaws through repeated tensile tests," Journal of the Royal Society Interface, 12(107), p. 20150177.

[111] Shah, D. U., Vollrath, F., Porter, D., Stires, J., and Deheyn, D. D., 2014, "Housing tubes from the marine worm Chaetopterus sp.: biomaterials with exceptionally broad

thermomechanical properties," Journal of the Royal Society Interface, 11(98), p. 20140525.

[112] Tarakanova, A., and Buehler, M. J., 2012, "The role of capture spiral silk properties in the diversification of orb webs," Journal of the Royal Society Interface, 9(77), pp. 3240-3248.

[113] Taylor, D., and Dirks, J.-H., 2012, "Shape optimization in exoskeletons and endoskeletons: a biomechanics analysis," Journal of the Royal Society Interface, 9(77), pp. 3480-3489.

[114] Wang, X., Ren, H., Zhang, B., Fei, B., and Burgert, I., 2011, "Cell wall structure and formation of maturing fibres of moso bamboo (Phyllostachys pubescens) increase buckling resistance," Journal of the Royal Society Interface, p. 20110462.

[115] Jacobs, P. S., and Rau, L. F., 1990, "SCISOR: Extracting information from on-line news," Communications of the ACM, 33(11), pp. 88-97.

[116] Wang, R., Suo, Z., Evans, A., Yao, N., and Aksay, I., 2001, "Deformation mechanisms in nacre," Journal of Materials Research, 16(09), pp. 2485-2493.

[117] Katti, K. S., Katti, D. R., Pradhan, S. M., and Bhosle, A., 2005, "Platelet interlocks are the key to toughness and strength in nacre," Journal of Materials Research, 20(05), pp. 1097-1100.

[118] Fu, Q., Saiz, E., and Tomsia, A. P., 2011, "Bioinspired strong and highly porous glass scaffolds," Advanced Functional Materials, 21(6), pp. 1058-1063.

[119] Mayer, G., 2011, "New toughening concepts for ceramic composites from rigid natural materials," Journal of the Mechanical Behavior of Biomedical Materials, 4(5), pp. 670-681.

[120] Wegst, U. G., 2011, "Bending efficiency through property gradients in bamboo, palm, and wood-based composites," Journal of the Mechanical Behavior of Biomedical Materials, 4(5), pp. 744-755.

[121] Chen, L., Ballarini, R., Kahn, H., and Heuer, A., 2007, "Bioinspired micro-composite structure," Journal of Materials Research, 22(01), pp. 124-131.

[122] Stanzl-Tschegg, S. E., Keunecke, D., and Tschegg, E. K., 2011, "Fracture tolerance of reaction wood (yew and spruce wood in the TR crack propagation system)," Journal of the Mechanical Behavior of Biomedical Materials, 4(5), pp. 688-698.

[123] Chan, Y., Ngan, A., and King, N., 2011, "Nano-scale structure and mechanical properties of the human dentine–enamel junction," Journal of the Mechanical Behavior of Biomedical Materials, 4(5), pp. 785-795.

[124] Erb, R. M., Libanori, R., Rothfuchs, N., and Studart, A. R., 2012, "Composites reinforced in three dimensions by using low magnetic fields," Science, 335(6065), pp. 199-204.

[125] Nardi, T., Leterrier, Y., Karimi, A., and Månson, J.-A. E., 2014, "A novel synthetic strategy for bioinspired functionally graded nanocomposites employing magnetic field gradients," RSC Advances, 4(14), pp. 7246-7255.

[126] Elices, M., Plaza, G. R., Pérez-Rigueiro, J., and Guinea, G. V., 2011, "The hidden link between supercontraction and mechanical behavior of spider silks," Journal of the Mechanical Behavior of Biomedical Materials, 4(5), pp. 658-669.

[127] Lin, T.-H., Huang, W.-H., Jun, I.-K., and Jiang, P., 2009, "Bioinspired assembly of colloidal nanoplatelets by electric field," Chemistry of Materials, 21(10), pp. 2039-2044.

[128] Vincent, J. F., 2011, "Unusual uses of holes—With input from biology," Journal of the Mechanical Behavior of Biomedical Materials, 4(5), pp. 682-687.

[129] Barthelat, F., 2010, "Nacre from mollusk shells: a model for high-performance structural materials," Bioinspiration & Biomimetics, 5(3), p. 035001.

[130] Zhang, Z., Liu, B., Zhang, Y.-W., Hwang, K.-C., and Gao, H., 2014, "Ultra-strong collagen-mimic carbon nanotube bundles," Carbon, 77, pp. 1040-1053.

[131] Caliari, S. R., Ramirez, M. A., and Harley, B. A., 2011, "The development of collagen-GAG scaffold-membrane composites for tendon tissue engineering," Biomaterials, 32(34), pp. 8990-8998.

[132] Aggarwal, C. C., and Zhai, C., 2012, "A survey of text clustering algorithms," *Mining Text Data*, Springer, New York, NY.

[133] Fung, B. C., Wang, K., and Ester, M., 2003, "Hierarchical document clustering using frequent itemsets," *SIAM International Conference on Data Mining*, San Francisco, CA.

## APPENDIX A

## STAGE 1 SEARCH TERMS

*terms added to the updated term sets

**Biological Terms**
biological
biology
biopolymer
biopolymers
biomaterial
biomaterials
bioscience
life
living
natural
nature
organic
organism
organisms
protein
proteins
tissue
tissues
phylum
genus
species
plant
plants
animal
animals

**Property Terms: Density**

*buoancy
*dense
*denser
densities
density
*lighter
mass per unit volume
*specific bending stiffness
specific volume
specific volumes
*to weight ratio
*weight
weight per unit volume

**Property Terms: Ductility**

brittle
brittleness
*deform
*deformable
ductile
ductility
elongation
elongations
*extensibility
*extensible
*failure length
*failure strain
fracture length
fracture lengths
fracture strain
fracture strains
original gauge length
original gauge lengths
percent reduction in area
percent reductions in area
*strain to failure
*strain-to-failure

**Property Terms: Hardness**

Barcol
Brinell
Durometer
hard
harden
*hardening
harder
hardest
hardness
hardnesses
HB
HR
indent
indentation
indentations
indenter
indenters
Knoop
microindentation
microindentations
Mohs
penetration
Rockwell
scratch
scratches
Shore
soft
*softening
softness
Vickers

**Property Terms: Stiffness**
*compliance
*compliant
*elastic deformation
*flexibility
*flexible
Hooke's law
moduli
modulus
*resistance to bending
*resistant to bending
rigidities
rigidity
stiff
stiffen
stiffer
stiffest
stiffness
*strain energy
Young's

**Property Terms: Strength**
*compressive force
*compressive forces
compressive strength
compressive strengths
*failure strength
*failure strengths
fracture strength
fracture strengths
*high load
*high loads
load carrying capacities
load carrying capacity
*load-bearing
*loaded in tension
maximum stress
maximum stresses
*strength
strengthen
strengthening
*strong
*tensile force
*tensile forces
tensile strength
tensile strengths
*theoretical strength
ultimate strength
ultimate strengths
yield point
yield points
yield strength
yield strengths
yield stress
yield stresses

**Property Terms: Toughness**
absorb energy
absorbs energy
crack propagation
crack propagations
*dissipate energy
*dissipates energy
energies absorbed
energy absorbed
energy absorption
*energy dissipated
*energy dissipation
failure mode
failure modes
*fracture energy
fracture resistance
fracture resistances
*impact resistance
*impact resistant
impact strength
impact strengths
impact test
impact tests
*impact-resistant
resist fracture
resistance to fracture
resistances to fracture
resists fracture
*strain energy
stress intensity factor
stress intensity factors
tough
toughen
toughened
*toughening
toughness

**Relationship Terms**

able to
account
accounted
accounting
accounts
act
acted
acting
acts
*affect
*affected
*affecting
*affects
allow
allowed
allowing
allows
*attribute
*attributed
*attributes
*attributing
because
by
cause
caused
causes
causing
change
changed
changes
changing
consequence
consequences
consequently
*consider
*considered
*considering
*considers
*construct
*constructed
*constructing
*construction
*constructions
*constructs
contribute
contributed
contributes
contributing
contribution
contributions
correspond
corresponded
corresponding
corresponds
decrease
decreased
decreases
decreasing
*demonstrate
*demonstrated
*demonstrates
*demonstrating
depend
depended
depending
depends
derive
derived
derives
deriving
determine
determined
determines
determining
due
effect
enable
enabled
enables
enabling
*examine
*examined
*examines
*examining
explain
explained
explaining
explains
explanation
explanations
for
force
forced
forces
forcing
from
function
functioned
functioning
functions
*gave
generate
generated
generates
generating
*give
*given
*gives
*giving
*impact
*impacts
*implication
*implications
*implied
*implies
*imply
*implying
in order
increase
increased
increases
increasing
indicate
indicated
indicates
indicating
influence
influenced
influences

influencing
*introduce
*introduced
*introduces
*introducing
*introduction
*introductions
lead
leading
leads
led
made
make
makes
making
maximize
maximized
maximizes
maximizing
mean
meaning
means
meant
minimize
minimized
minimizes
minimizing
*observation
*observations
*observe
*observed
*observes
*observing
*offer
*offered
*offering
*offerings
*offers
provide
provided
provides
providing
purpose
purposes
reason
reasoned
reasoning
reasons
reduce
reduced
reduces
reducing
relate
related
relates
relating
*resolution
*resolutions
*resolve
*resolved
*resolves
*resolving
result
resulted
resulting
results
*reveal
*revealed
*revealing
*reveals
role
roles
*show
*showed
*showing
*shown
*shows
so
spread
spreading
spreads
*suggest
*suggested
*suggesting
*suggestion
*suggestions
*suggests
*tend
*tended
*tending
*tends
therefore
thus
*through
*use
*used
*uses
*using

# APPENDIX B

## STAGE 3 IRRELEVANT PASSAGE PATTERN TERM SETS

**Updated Property Terms: Density**

buoyancy
dense
denser
densest
densities
density
lightening
specific bending stiffness
specific volume
specific volumes
weight
weights

**Updated Property Terms: Ductility**

brittle
brittlely
brittleness
brittler
brittlest
ductile
ductilely
ductilities
ductility
extensibilities
extensibility
extensible
failure length
failure lengths
failure strain
failure strains
fracture length
fracture lengths
fracture strain
fracture strains
original gauge length
original gauge lengths
percent reduction in area
percent reduction in areas
percent reductions in area
strain to failure
strain to fracture
strains to failure
strains to fracture

**Updated Property Terms: Hardness**

hard
harden
hardening
harder
hardest
hardness
hardnesses
indentation map
indentation maps
microindentation map
microindentation maps
nanoindentation map
nanoindentation maps
soft
soften
softening
softer
softest
softness
softnesses

**Updated Property Terms: Stiffness**
elastic deformation
elastic deformations
hookes law
Hooke's law
moduli
modulus
resistance to bending
resistances to bending
resistant to bending
rigid
rigidities
rigidity
rigidly
rigidness
stiff
stiffen
stiffened
stiffening
stiffer
stiffest
stiffly
stiffness
stiffnesses
strain energy

**Updated Property Terms: Strength**
load carrying capacities
load carrying capacity
modulus of rupture
strength
strengthen
strengthener
strengthening
strengths
stress
stresses
stronger
tensile force
tensile forces
yield point
yield points

**Updated Property Terms: Toughness**
absorb energy
absorbing energies
absorbing energy
absorbs energy
absorption of energy
crack propagation
crack propagations
dissipate energies
dissipate energy
dissipated energy
dissipates energy
dissipation of energy
energies absorbed
energy absorbed
energy absorption
energy dissipated
energy dissipation
fracture energies
fracture energy
fracture resistance
fracture resistances
fracture resistant
impact resistance
impact resistances
impact resistant
impact strength
impact strengths
resist fracture
resistance to fracture
resistances to fracture
resistant to fracture
resists fracture
strain energies
strain energy
tough
toughen
toughened
toughening
tougher
toughest
toughness
toughnesses

**Exclusion Terms:**
**Density**

absorption density
anion densities
areal densities
areal density
atom density
atomic densities
atomic density
atomic number densities
atomic weight
atomic weights
band densities
band density
bit densities
bit density
branch densities
branch density
by weight
calorie dense
capacitance densities
capacitance density
carrier densities
carrier density
cation density
cell densities
cell density
chain densities
chain density
charge densities
charge density
cluster densities
cluster density
crack densities
crack density
crosslink densities
crosslink density
crosslinking densities
crosslinking density
crystallographic densities
current densities
current density
data density
defect dense
defect densities
defect density
dense arms
dense array
dense arrays
dense bend contours
dense cell
dense concentration
dense crystal packing
dense defects
dense DNA packing
dense fibrillar matrix
dense flow
dense flows
dense framework
dense frameworks
dense grains
dense granular flows
dense group
dense groups
dense inclined flows
dense integration
dense lattice
dense limit
dense mesh
dense meshes
dense molecular packing
dense nanotube arrays
dense network
dense nucleation
dense packing
dense packings
dense parallel arrays
dense pattern
dense patterns
dense periodic
dense plane
dense planes
dense population
dense populations
dense random
dense regular packing
dense spots
denser arrays
denser concentration
denser grid
denser nucleation
denser packing
denser plane
densest arrays
densest concentration
densest packing
densest plane
densities of cells
densities of charge
densities of impurities
densities of state
densities of states
density array
density arrays
density carriers
density dislocations
density driven
density functional
density of carriers
density of charge
density of clusters
density of crosslinks
density of defects
density of dipoles
density of dislocations
density of hairs
density of particles
density of point defects
density of pores
density of power
density of state
density of states
density of surface defects
density of the dislocations

density packing
density relaxation
density storage
density waves
deposition densities
dipole density
dislocation densities
dislocation density
DNA densities
dopant densities
dopant density
doping densities
dot densities
drop weight
dry weight
electrode density
electron dense
electron densities
electron density
electronic densities
electronic density
electronically denser
energy dense
energy densities
energy density
entanglement densities
entanglement density
excitation density
fault densities
fault density
flux densities
flux density
framework densities
framework density
graft densities
graft density
grafting densities
grid densities
grid density
hairpin densities
hairpin density
hole densities
hole density
image density
information density
integration densities
interlink densities
interlink density
junction densities
junction density
length densities
ligand densities
ligand density
local density
approximation
magnetization densities
magnetization density
memory densities
memory density
mineral densities
mineral density
molecular densities
molecular density
molecular weight
molecular weights
momentum density
nucleation densities
nucleation density
nucleus density
number density
optical densities
optical density
ordering densities
packaging densities
packed densities
packed density
packing densities
packing density
pair densities
particle densities
particle density
particle number
densities
photocurrent densities
photocurrent density
pinning densities
pinning density
probability weight
polarization densities
polarization density
polymer weight ratio
polymer weight ratios
population densities
population density
pore densities
pore density
power dense
power densities
power density
precipitate densities
precipitate density
probability densities
probability density
receptor densities
receptor density
recording densities
recording density
seeding densities
seeding density
segment densities
segment density
site density
spatial densities
spatial density
spectral densities
spectral density
spectral weight
spin densities
spin density
stack densities
stack density
stacking densities
stacking density
storage densities
storage density
supercurrent density
surface densities
surface density
thread densities
thread density
transistor density
vortex density

weight average
weight averaged
weight bearing
weight concentration
weight concentrations
weight distribution
weight distributions
weight fraction
weight fractions
weight percent
weight percentage
weight ratio
weight ratios
weight sensitivity
weight sensitivities

**Exclusion Terms:**
**Ductility**
brittle star
brittle stars
fatigue ductility

**Exclusion Terms:**
**Hardness**

beam hardening
cyclic hardening
hard axes
hard axis
hard choice
hard choices
hard contact
hard contacts
hard core
hard disk
hard disks
hard drive
hard drives
hard indenter
hard light
hard matter
hard question
hard questions
hard sphere
hard spheres
hard tapping
hard template
hard templates
hard templating
hard thing
hard tip
hard to
hard way
hard wearing
hard x ray
hard x rays
harder contact
harder contacts
harder for
harder than
harder to
hardness h
hardness hk
hardness hv
hardness indentation
hardness indentations
hardness indenter
hardness indent
hardness indents
hardness scale
hardness tester
indentation hardness
knoop hardness
magnetically soft
microhardness
precipitation hardening
radiation hardening
soft adhesion
soft bake
soft baked
soft chemistry
soft cloth
soft contact
soft contacts
soft elasticity
soft eye
soft imaging
soft knee
soft landed
soft landing
soft lithographic
soft lithography
soft magnetic
soft magnets
soft matter
soft method
soft mode
soft molding
soft photolithography
soft science
soft sphere
soft system
soft systems
soft tapping
soft technique
soft template
soft templates
soft templating
soft x ray
soft x rays
softening temperature
solid solution hardening
strain hardening
strain rate hardening
strain softening
stress hardening
vickers hardness
water hardness
work hardening
work softening

**Exclusion Terms:**
**Stiffness**

assume elastic deformation
bending modulus eb
bulk modulus b
bulk modulus k
by means of hookes law
chain rigidity
complex moduli
complex modulus
contact moduli
contact modulus
contact stiffness
dynamic moduli
dynamic modulus
elastic deformation energy
elastic deformation fields
elastic modulus e
elimination modulus
equilibrium elastic deformation
exchange stiffness
fourier moduli
fourier modulus
gaussian bending rigidity
indentation moduli
indentation modulus
loss moduli
loss modulus
magnetic stiffness
moduli data
modulus g
modulus measurements
modulus of capacitance
modulus of toughness
molecular rigidity
plateau moduli
plateau modulus
purely elastic deformations
range of moduli
repeated elastic deformation
rigid backbone
rigid body
rigid building
rigid chain
rigid groups
rigid molecular
rigid molecules
rigid rod
rigid rotation
rigid rotations
rigid shift
rigidly fixed
rigidly shearing
rubber modulus
rubbery moduli
rubbery modulus
shear modulus g
stiff indenter
storage moduli
storage modulus
stretch moduli
stretch modulus
tilt moduli
tilt modulus
tilt stiffness
viscous moduli
viscous modulus
weibull modulus
youngs modulus e
youngs modulus y

**Exclusion Terms:**
**Strength**

absorption strength
actuation stress
adhesion strength
adhesion strengths
adhesive strength
adhesive strengths
after the yield point
applied stress
applied stresses
before the yield point
beyond the yield point
binding strength
binding strengths
bond strength
bond strengths
bonding strength
bonding strengths
chiral strength
clamping stress
compatibility stresses
constant stress
contact strength
coupling strength
coupling strengths
creep strengthening
crystal strength
cyclic stress
cyclic stresses
dielectric strength
electrostatic stress
epitaxial stress
exchange strength
fatigue strength
fatigue strengths
fatigue stress
fatigue stresses
field strength
field strengths
filter strength
flow stress
frictional strength
full strength
gel strength
gel strengths
generated stress
hybridization strength
impact strength
impact strengths
important to stress
initiation stress
interaction strength
interaction strengths
interfacial strength
interfacial strengths
interfacial stress
interfacial stresses
intergranular stress
intergranular stresses
interlaminar strength
internal stress
internal stresses
ionic strength
ionic strengths
key strength
major strength of
nucleation stress
optical strength
oscillator strength
oscillator strengths
osmotic stress
peierls stress
periodic stress
pull off strength
pull off strengths
readings of stress
relaxation of stresses
relaxation strength
relaxation strengths
relaxes stresses
residual stress
residual stresses
shear stress tau
signal strengths
strength of adhesion
strength of chirality
strength of interaction
strength of magnetic
strength of magnetization
strength of superconductivity
strength of the
strength of these
strength tester
strengths and weaknesses
strengths in
stress amplitude
stress corrosion
stress free
stress function
stress generation
stress intensities
stress intensity
stress plot
stress range
stress rate
stress reading
stress recovery
stress relaxation
stress relaxes
stress sigma
stress strain
stress tensors
stress that
stress the importance
stress whitening
stresses of
stresses the importance
stronger approach
stronger attraction
stronger attractive
stronger binding
stronger bonding
stronger dependence
stronger dipoles
stronger impact
stronger increase

stronger influence
stronger interaction
stronger interfacial
stronger tendency
stronger volume
stronger wear
surface stress
surface stresses
the stronger the
to stress
transition strength
transition strengths
unit of stress
up to a yield point
viscous stresses
which stresses how
zero stress

**Exclusion Terms:**
**Toughness**

crack propagation rate
dissipated energy density
dissipation of energy flux
energy absorption band
energy absorption peak
energy dissipation density
indentation toughness
interface fracture energies
interface fracture energy
interface fracture resistance
interface fracture resistances
interface fracture toughness
interface toughness
interfacial fracture energies
interfacial fracture energy
interfacial fracture resistance
interfacial fracture resistances
interfacial fracture toughness
interfacial toughness
interlaminar fracture toughness
ion energy dissipation
measurement of toughness
molecular energy dissipation
rate of crack propagation
strain energy function
tackle tough
tough competition
tough decision
tough decisions
tough game
tough one
tough problem
tough problems
tough to
toughening agent
toughening agents
tougher criteria
toughness gic
toughness giic
toughness kc
toughness kic
toughness measurement
toughness measurements
toughness parameter
toughness testing
toughness value
toughness values
velocity toughening

## APPENDIX C

## STAGE 3 RELEVANT PASSAGE PATTERN TERM SETS

**Stand-Alone Terms: Density**
denser
densest
densified
lightening
lightweight

**Stand-Alone Terms: Ductility**
(none)

**Stand-Alone Terms: Hardness**
harden
hardened
hardening
hardens
harder
hardest
soften
softened
softening
softens
softer
softest

**Stand-Alone Terms: Stiffness**
resistant to bending
stiffen
stiffening
stiffer
stiffest

**Stand-Alone Terms: Strength**
strengthen
strengthened
strengthening
stronger
strongest
weakened
weakens
weaker
weakest

**Stand-Alone Terms: Toughness**
crack propagation
toughen
toughened
toughening
toughens
tougher
toughest
toughness

**Identifier Terms**

affect
affected
affecting
affects
attribute
attributed
attributes
attributing
cause
caused
causes
causing
change
changed
changes
changing
close
closer
closest
comparable
correlate
correlated
correlates
correlating
correlation
correlations
decrease
decreased
decreases
decreasing
define
defined
defines
defining
degree
depend
depended
dependence
dependent
depending
depends
differ
differed
difference
differences
different
differing
differs
due
effect
effects
enhance
enhanced
enhancement
enhancements
enhances
enhancing
equal
equaled
equalled
equaling
equalling
equals
excellent
excellently
exceptional
exceptionally
few
fewer
fewest
function
functioned
functioning
functions
gave
give
gives
giving
good
great
greater
greatest
half
high
higher
highest
highly
hinder
hindered
hindering
hinders
huge
improve
improved
improvement
improves
improving
increase
increased
increases
increasing
incremental
influence
influenced
influences
influencing
large
larger
largest
lead
leading
leads
least
led
less
low
lower
lowered
lowering
lowers
lowest
maximum
moderate
moderately
more
most
outstanding
outstandingly
presence

presences
prone
propose
proposed
proposes
proposing
range
ranges
reduce
reduced
reduces
reducing
reduction
relation
relatively
result
resulted
resulting
results
scale
significant
significantly
similar
similarly
small
smaller
smallest
some
strong
stronger
strongest
strongly
tailor
tailored
tailoring
tailors
twice
variation
variations
varied
varies
vary
varying

## APPENDIX D

## REVISED STAGE 1 ALGORITHM CODE

### BioinspiredMaterialsMain Class

```
import java.io.File;
import java.io.FileWriter;
import java.io.IOException;
import java.io.PrintWriter;
import java.util.ArrayList;
import java.util.Arrays;
import java.util.Scanner;

public class BioinspiredMaterialsMain {

	private static final String[] strengthPhrases = {"compressive strength", "compressive
strengths", "fracture strength", "fracture strengths", "load carrying capacities", "load carrying capacity",
"maximum stress", "maximum stresses", "strengthen", "strengthening", "tensile strength", "tensile
strengths", "ultimate strength", "ultimate strengths", "yield point", "yield points", "yield strength", "yield
strengths", "yield stress", "yield stresses", "strong", "strength", "load bearing", "loaded in tension", "high
load", "high loads", "failure strength", "failure strengths", "theoretical strength", "tensile force", "tensile
forces", "compressive force", "compressive forces"};
	private static final String[] ductilityPhrases = {"brittle", "brittleness", "ductile", "ductility",
"elongation", "fracture length", "fracture strain", "original gauge length", "percent reduction in area",
"elongations", "fracture lengths", "fracture strains", "original gauge lengths", "percent reductions in area",
"strain to failure", "strain to failure", "failure strain", "failure length", "deformable", "deform",
"extensible", "extensibility"};
	private static final String[] stiffnessPhrases = {"hooke's law", "moduli", "modulus", "rigidity",
"stiff", "stiffer", "stiffen", "stiffest", "stiffness", "young's", "rigidities", "strain energy", "elastic
deformation", "resistant to bending", "resistance to bending", "flexible", "flexibility", "compliant",
"compliance"};
	private static final String[] toughnessPhrases = {"absorb energy", "absorbs energy", "crack
propagation", "crack propagations", "energies absorbed", "energy absorbed", "energy absorption", "failure
mode", "failure modes", "fracture resistance", "fracture resistances", "impact strength", "impact strengths",
"impact test", "impact tests", "resist fracture", "resistance to fracture", "resistances to fracture", "resists
fracture", "stress intensity factor", "stress intensity factors", "tough", "toughen", "toughened", "toughness",
"toughening", "strain energy", "impact resistant", "impact resistance", "impact-resistant", "fracture
energy", "energy dissipation", "energy dissipated", "dissipates energy", "dissipate energy"};
	private static final String[] hardnessPhrases = {"barcol", "brinell", "durometer", "hard",
"harder", "hardest", "harden", "hardness", "hb", "hr", "indenter", "indent", "indentation", "knoop",
"microindentation", "mohs", "penetration", "rockwell", "scratch", "shore", "soft", "softness", "vickers",
"hardnesses", "indenters", "indentations", "microindentations", "scratches", "hardening", "softening"};
	private static final String[] densityPhrases = {"densities", "density", "mass per unit volume",
"specific volume", "weight per unit volume", "specific volumes", "weight", "specific bending stiffness",
"buoancy", "to weight ratio", "lighter"};

	public static void main(String[] args) throws IOException
	{
		long startTime = System.currentTimeMillis();
		Scanner kbInput = new Scanner(System.in);
```

```
		{
			System.out.println("!!PLEASE MAKE SURE TO READ THE READ-ME
FILE BEFORE USING THE PROGRAM!!");
			System.out.println("Please enter number(s) of the desired property: [125]");
			System.out.println("1. strength\n2. ductility\n3. stiffness\n4. toughness\n" +
						"5. hardness\n7. density");
			String choices = kbInput.nextLine();
			ArrayList<ArrayList<String>> searchTermLists = selectSearchTerms(choices);

			String fileName = getFilename(choices);
			PrintWriter writer = new PrintWriter(new FileWriter(fileName, false));

			File folder = new File("./journals");
			for(String s : folder.list())
				System.out.println(s);

			BioLibrary library;
			for(File file: folder.listFiles())
			{
				library = new BioLibrary(file, searchTermLists);
				String result = library.searchArticles();
				if(result.equals(""))
				{
					result = "No result found for the property(s) selected";
					System.out.println(result);
				}
				else
					System.out.println(result);
				System.out.println("Number of Articles searched in " + file + " is " +
library.getNumberOfArticlesSearched());

				writer = new PrintWriter(new FileWriter(fileName, true));
				writer.println(result);
				writer.close();
			}
		}
		long endTime = System.currentTimeMillis();
		System.out.println("Search took " + (endTime - startTime) + " milliseconds");
	}

	private static String getFilename(String s)
	{
		String temp = "";
		for(int i = 0; i < s.length(); i++)
		{
			char c = s.charAt(i);
			switch(c-48)
			{
				case 1:
					temp += "strength";
					break;
				case 2:
```

```
						temp += "ductility";
						break;
					case 3:
						temp += "stiffness";
						break;
					case 4:
						temp += "toughness";
						break;
					case 5:
						temp += "hardness";
						break;
					case 7:
						temp += "density";
						break;
				}
				if(!(i==s.length()-1))
					temp += " ";
			}
			temp += ".txt";
			return temp;
		}

		private static ArrayList<ArrayList<String>> selectSearchTerms(String s)
		{
			ArrayList<ArrayList<String>> temp = new ArrayList<ArrayList<String>>();
			for(char c : s.toCharArray())
			{
				switch(c-48)
				{
					case 1:
						temp.add(new
ArrayList<String>(Arrays.asList(strengthPhrases)));
						System.out.println(c);
						break;
					case 2:
						temp.add(new
ArrayList<String>(Arrays.asList(ductilityPhrases)));
						System.out.println(c);
						break;
					case 3:
						temp.add(new
ArrayList<String>(Arrays.asList(stiffnessPhrases)));
						System.out.println(c);
						break;
					case 4:
						temp.add(new
ArrayList<String>(Arrays.asList(toughnessPhrases)));
						System.out.println(c);
						break;
					case 5:
						temp.add(new
ArrayList<String>(Arrays.asList(hardnessPhrases)));
						System.out.println(c);
```

```
					break;
				case 7:
					temp.add(new
ArrayList<String>(Arrays.asList(densityPhrases)));
					System.out.println(c);
					break;
			}
		}
		return temp;
	}
}
```

### BioLibrary Class

```
import java.io.BufferedReader;
import java.io.File;
import java.io.FileReader;
import java.io.IOException;
import java.util.ArrayList;

//BioLibrary class - takes the searchTermLists from main, opens a text file and searches through the
articles
//responsible for:
// 1. parsing through the article for sentences and create Article object
// 2. calling the search algorithm from Article

public class BioLibrary {

	private BufferedReader fileInput;
	private Article currentArticle;

	private ArrayList<ArrayList<String>> searchTermLists;
	private String searchResult;

	private boolean getInfo;
	private boolean getContent;

	private int numArticles;

	public BioLibrary(File file, ArrayList<ArrayList<String>> list) throws IOException
	{
		fileInput = new BufferedReader(new FileReader(file));
		searchTermLists = list;

		currentArticle = null;
		searchResult = "";
		getInfo = false;
		getContent = false;

		findFirstArticle();
	}
```

```
private void findFirstArticle() throws IOException
{
	String line = "";

	while((line = fileInput.readLine()) != null)
	{
		if(line.equals("##"))
			break;
	}

	getInfo = true;
	getContent = false;
}

private void loadNext() throws IOException
{
	if(!getInfo)
	{
		currentArticle = null;
		return;
	}

	String line = "";
	String info = "";

	Sentence tempSentence = new Sentence();

	while((line = fileInput.readLine()) != null)
	{
		if(line.equals("##"))
		{
			getContent = false;
			getInfo = true;
			info = "";
			break;
		}

		if(getInfo)
		{
			if(line.equals("#"))
			{
				getInfo = false;
				getContent = true;
				currentArticle = new Article(info);
			}
			else
			{
				line = line.trim();
				if(!line.equals(""))
				{
					info += line;
					info += "\n";
				}
```

```
                }
                continue;
            }

            if(getContent)
            {
                line = line.trim();
                for(String s : line.split("\\s+|-"))
                {
                    if(s.equals(""))
                        continue;

                    s = s.replaceAll("[F|f]ig\\.", "fig");
                    s = s.replaceAll("[E|e]q\\.", "eq");
                    s = s.replaceAll("al\\.", "al");
                    s = s.replaceAll("etc\\.", "etc");
                    s = s.replaceAll("i\\.e\\.", "ie");
                    s = s.replaceAll("e\\.g\\.", "eg");
                    s = s.replaceAll("\\[.+\\]", "");
                    s = s.replaceAll("[^a-zA-Z0-9_.?!%/\'\"]", "");

                    if(s.equals(""))
                        continue;
                    char lastChar = s.charAt(s.length()-1);

                    s = s.replaceAll("[.?!]", "");
                    if(s.equals(""))
                        continue;

                    if(lastChar == '.' || lastChar == '?' || lastChar == '!')
                    {
                        tempSentence.addWord(s.replaceAll("[.?!]", ""));
                        if(tempSentence.length() > 2)
                            currentArticle.addSentence(tempSentence);
                        tempSentence = new Sentence();
                    }
                    else
                    {
                        tempSentence.addWord(s);
                    }
                }
            }
        }

        if(!tempSentence.toString().equals(""))
            currentArticle.addSentence(tempSentence);
    }

    public String searchArticles() throws IOException
    {
        loadNext();
        while(currentArticle != null)
        {
```

```
				numArticles++;
				if(currentArticle.isBioRelated())
				{
					String temp = currentArticle.runSearchAlgorithm(searchTermLists);

					if(!temp.equals(""))
					{
						searchResult += "Article Info:\n" + currentArticle.getInfo();
						searchResult += "Search Result:\n" + temp + "\n\n";
					}
				}
				loadNext();
			}
			return searchResult;
		}

		public int getNumberOfArticlesSearched()
		{
			return numArticles;
		}
}
```

## Article Class

```
import java.util.ArrayList;
import java.util.Arrays;
import java.util.HashSet;
import java.util.TreeSet;

//Article class - saves the biographical info and sentences of an article
// responsible for running searching algorithm as well as static variables including:
// "bioPhrases", "relationshipPhrases" and parameters for searching range

public class Article {

		private static final String[] bioPhrases = {"biological", "biology", "biopolymer", "biopolymers",

		"biomaterial", "biomaterials", "bioscience", "life",

		"living", "natural", "nature", "organic", "organism",

		"organisms", "protein", "proteins", "tissue", "tissues",

		"phylum", "genus", "species", "plant", "plants", "animal",

		"animals"};
		private static final String[] relationshipPhrases = {"able to", "account", "accounted",
"accounting", "accounts",
				"act", "acted", "acting", "acts", "affect", "affected", "affecting", "affects", "allow",
"allowed", "allowing",
				"allows", "attribute", "attributed", "attributes", "attributing", "because", "by", "cause",
"caused", "causes",
```

```
			"causing", "change", "changed", "changes", "changing", "consequence",
"consequences", "consequently", "consider",
			"considered", "considering", "considers", "construct", "constructed", "constructing",
"construction",
			"constructions", "constructs", "contribute", "contributed", "contributes", "contributing",
"contribution",
			"contributions", "correspond", "corresponded", "corresponding", "corresponds",
"decrease", "decreased",
			"decreases", "decreasing", "demonstrate", "demonstrated", "demonstrates",
"demonstrating", "depend", "depended",
			"depending", "depends", "derive", "derived", "derives", "deriving", "determine",
"determined", "determines",
			"determining", "due", "effect", "enable", "enabled", "enables", "enabling", "examine",
"examined", "examines",
			"examining", "explain", "explained", "explaining", "explains", "explanation",
"explanations", "for", "force",
			"forced", "forces", "forcing", "from", "function", "functioned", "functioning",
"functions", "gave", "generate",
			"generated", "generates", "generating", "give", "given", "gives", "giving", "impact",
"impacts", "implication",
			"implications", "implied", "implies", "imply", "implying", "in order", "increase",
"increased", "increases",
			"increasing", "indicate", "indicated", "indicates", "indicating", "influence", "influenced",
"influences",
			"influencing", "introduce", "introduced", "introduces", "introducing", "introduction",
"introductions", "lead",
			"leading", "leads", "led", "made", "make", "makes", "making", "maximize",
"maximized", "maximizes", "maximizing",
			"mean", "meaning", "means", "meant", "minimize", "minimized", "minimizes",
"minimizing", "observation",
			"observations", "observe", "observed", "observes", "observing", "offer", "offered",
"offering", "offerings",
			"offers", "provide", "provided", "provides", "providing", "purpose", "purposes",
"reason", "reasoned", "reasoning",
			"reasons", "reduce", "reduced", "reduces", "reducing", "relate", "related", "relates",
"relating", "resolution",
			"resolutions", "resolve", "resolved", "resolves", "resolving", "result", "resulted",
"resulting", "results", "reveal",
			"revealed", "revealing", "reveals", "role", "roles", "show", "showed", "showing",
"shown", "shows", "so", "spread",
			"spreading", "spreads", "suggest", "suggested", "suggesting", "suggestion",
"suggestions", "suggests", "tend", "tended",
			"tending", "tends", "therefore", "thus", "through", "use", "uses", "used", "using"};

	private static final int relationshipRange = 0;

	private String info;
	private boolean bioRelated;

	private ArrayList<Sentence> sentences;

	public Article(String i)
	{
```

```
		info = i;
		bioRelated = false;
		sentences = new ArrayList<Sentence>();
	}

	public void addSentence(Sentence s)
	{
		sentences.add(s);
		if(!bioRelated && s.matchPhrases(new ArrayList<String>(Arrays.asList(bioPhrases))))
		{
			bioRelated = true;
		}
	}

	public boolean isBioRelated()
	{
		return bioRelated;
	}

	public String getInfo()
	{
		return info;
	}

	public String toString()
	{
		String content = "";
		for(Sentence s : sentences)
		{
			content += s.toString();
			content += "\n";
		}
		return "Article Info:\n" + info + "\nContent:\n" + content;
	}

	public String runSearchAlgorithm(ArrayList<ArrayList<String>> searchTermList)
	{
		ArrayList<HashSet<Integer>> matchingIndexes =
getPropertyMatchingIndexes(searchTermList);
		HashSet<Integer> mergedIndexes =
getMergedIndexes(propertyHash(matchingIndexes), getRelationshipIndexes());
		TreeSet<Integer> returnIndexes = new TreeSet<Integer>(mergedIndexes);
		String returnSentences = "";
		for(Integer i : returnIndexes)
		{
			if (i-1>=0)
				returnSentences += sentences.get(i-1) + "\n";
			returnSentences += sentences.get(i) + "\n";
			if (i+1<sentences.size())
				returnSentences += sentences.get(i+1) + "\n";
			returnSentences += "\n";
		}
		return returnSentences;
```

```
}

//STEP 1:
//for each property:
//1. take an ArrayList of property search terms for each property
//2. return an ArrayList of the indexes of the sentences that have search terms
private ArrayList<HashSet<Integer>>
getPropertyMatchingIndexes(ArrayList<ArrayList<String>> searchTermList)
{
	ArrayList<HashSet<Integer>> allMatchingIndexes = new
ArrayList<HashSet<Integer>> ();
	for(ArrayList<String> searchTerms : searchTermList)
	{
		allMatchingIndexes.add(getMatchingIndexes(searchTerms));
	}
	return allMatchingIndexes;
}

//STEP 2:
//identifies indexes of sentences containing all properties selected
private HashSet<Integer> propertyHash(ArrayList<HashSet<Integer>> matchingIndexes)
{
	HashSet<Integer> combinedIndexes = new HashSet<Integer> (matchingIndexes.get(0));
	for(int i = 1; i < matchingIndexes.size(); i++)
	{
		HashSet<Integer> tempIndexes = new HashSet<Integer>
(matchingIndexes.get(i));
		combinedIndexes.retainAll(tempIndexes);
	}
	return combinedIndexes;
}

//STEP 3:
//finds the index of sentences containing relationship keywords + "relationshipRange" number of
sentences before and after
//returns the indexes of the acceptable range
private HashSet<Integer> getRelationshipIndexes()
{
	HashSet<Integer> returnIndexes = new HashSet<Integer> ();
	for(Integer matchIndex : getMatchingIndexes(new
ArrayList<String>(Arrays.asList(relationshipPhrases))))
	{
		for(int j = 0; j <= relationshipRange; j++)
		{
			returnIndexes.add(matchIndex);
		}
	}
	return returnIndexes;
}

//STEP 4:
//take intersection of property indexes and relationship indexes
```

```
	private HashSet<Integer> getMergedIndexes(HashSet<Integer> propertyIndexes,
HashSet<Integer> relationshipIndexes)
	{
		HashSet<Integer> returnIndexes = new HashSet<Integer>(propertyIndexes);
		returnIndexes.retainAll(relationshipIndexes);
		return returnIndexes;
	}

	//helper function:
	//take an ArrayList of search terms and return the matching indexes
	private HashSet<Integer> getMatchingIndexes(ArrayList<String> searchTerms)
	{
		HashSet<Integer> matchingIndexes = new HashSet<Integer> ();
		for(int index = 0; index < sentences.size(); index++)
		{
			if(sentences.get(index).matchPhrases(searchTerms))
			{
					matchingIndexes.add(index);
			}
		}
		return matchingIndexes;
	}
}
```

## Sentence Class

```
import java.util.ArrayList;
import java.util.HashSet;

//Sentence class - saves each word in a sentence in a HashSet and an ArrayList

public class Sentence {
	private ArrayList<String> wordList;
	private HashSet<String> wordHash;

	public Sentence()
	{
		wordList = new ArrayList<String>();
		wordHash = new HashSet<String>();
	}

	public void addWord(String word)
	{
		wordList.add(word.toLowerCase());
		wordHash.add(word.toLowerCase());
	}

	public boolean matchPhrases(ArrayList<String> phrases)
	{
		String s = getSentence();
		for(String phrase : phrases)
		{
```

```
				if (wordHash.contains(phrase) || (phrase.contains(" ") && s.contains(phrase)))
				{
					return true;
				}
			}
			return false;
		}

		public String getSentence()
		{
			return wordList.toString().replaceAll("[\\[\\],]", "");
		}

		public String toString()
		{
			return getSentence();
		}

		public int length()
		{
			return wordList.size();
		}
}
```

# APPENDIX E

## STAGE 4 IRRELEVANT PASSAGE PATTERN TERM SETS

| **Method Terms** | **Calculation Terms** | **Result Terms** |
|---|---|---|
| a constant | area under | analytical method |
| crosshead | calculated | analytical methods |
| cycle | compute | assumed |
| cycles | defined as | curve |
| experimental | determined by | curves |
| extensometer | determined for | data |
| frequencies | parameter | slope |
| frequency | parameters | slopes |
| indentation load | taken as | values |
| instrumented | taken to be | |
| load application | e is the elastic modulus | |
| loading unloading | energy g | |
| measured | energy gf | |
| measurement | hardness h | |
| measurements | length l | |
| mold | load fmax | |
| molds | mass m | |
| rate | modulus e | |
| rates | modulus et | |
| sampled from | rigidity ei | |
| setup | stiffness e | |
| specimen | stiffness ei | |
| specimens | strength yf | |
| test | stress y | |
| tested | y the yield strength | |
| testing | | |
| tests | | |

# APPENDIX F

## STAGE 4 RELEVANT PASSAGE PATTERN TERM SETS

**Modification Terms**

a measure of
adding to
adjusting
alter
considerable
decrease
decreased
decreases
decreasing
degree of
enhanced
excellent
exceptional
exceptionally
exponential
full
gain
good
greater
high
higher
highest
improve
improved
improving
increase
increased
increases
increasing
large
less
low
lower
lowered
lowers
lowest
maximum
minimizes
minimum
more
most
negative
optimized
outstanding
per cent
percent
reduce
reduced
reduces
reducing
reduction
reinforcement
same?
significant
similar
smaller
sufficient
times as
twice as
variation
varying

**Causation Terms**

ability of
ability to
achieved
achieving
act as
acts as
acts to
affecting
affects
allows for
ascribed to
associated
attributable to
attributed
because
by
cause a
caused
causing
connected to
consequence of
consequently
contribute to
contributes to
contributing to
contribution
convey
correlate
correspond to
corresponds to
demonstrated
depend on
dependences
depending on
depends on
due to
effect
effects
enable
evidence
exhibited
explain
explained
explains
function of
gave
give
gives
imparts
implying that
in order to
inducing
influence of
lead to
leading
leads to
led to
linked to
makes
making
means to
modifies
produces
producing
promotes
provide
provoke
reason
related
relation
responsible for
result in
result of
resulted in
resulting from
resulting in
results in
satisfies
shown to
shows how
since
thereby
therefore
thus
to affect
to cause
to control
to induce
to make
to produce
to promote
to yield
used to
which
why
with
yielded
yielding
yields
if….then statements

**Correlation Terms**
correlated to
correlation
correlations
corresponding
dependence of
dependence on
dependent on
govern the
influence on
link between
provides
providing
role in
role of
roles in
ruled by

**Comparison Terms**
changed from
compared to
difference between
in comparison
in contrast to
in contrast with
than
the other hand
when comparing
whereas
while

# APPENDIX G

## STAGE 4 PART-OF-SPEECH TAGGING WITH FULL TAGSET

### Causation Terms

| Passage | IN | IN NN TO | IN TO | JJ TO | NN | NN IN | NN TO | NNS | RB | TO | TO VB | VB | VB DT | VB IN | VB TO | VBD | VBD IN | VBG | VBG IN | VBN | VBN TO | VBP | VBP TO | VBZ | VBZ TO | VBZ WRB | VBZ IN | VBZ TO | WDT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Relevant | 1.09 | 0.02 | 0.02 | 0.10 | 0.26 | 0.07 | 0.02 | 0.05 | 0.10 | 0.00 | 0.03 | 0.03 | 0.02 | 0.02 | 0.05 | 0.07 | 0.02 | 0.05 | 0.02 | 0.02 | 0.02 | 0.00 | 0.02 | 0.03 | 0.00 | 0.00 | 0.05 | 0.02 | 0.29 |
| Irrelevant | 1.27 | 0.02 | 0.00 | 0.10 | 0.21 | 0.02 | 0.03 | 0.06 | 0.13 | 0.02 | 0.02 | 0.03 | 0.00 | 0.01 | 0.02 | 0.01 | 0.00 | 0.01 | 0.02 | 0.06 | 0.02 | 0.03 | 0.00 | 0.02 | 0.01 | 0.01 | 0.01 | 0.01 | 0.24 |

### Correlation Terms

| Passage | NN | NN IN | NNS | VBG | VBG TO | VBP DT | VBZ |
|---|---|---|---|---|---|---|---|
| Relevant | 0.16 | 0.07 | 0.02 | 0.00 | 0.02 | 0.02 | 0.02 |
| Irrelevant | 0.00 | 0.02 | 0.00 | 0.01 | 0.00 | 0.00 | 0.01 |

### Comparison Terms

| Passage | DT JJ NN | IN | IN NN | NN IN | VBN TO |
|---|---|---|---|---|---|
| Relevant | 0.00 | 0.52 | 0.02 | 0.00 | 0.00 |
| Irrelevant | 0.01 | 0.40 | 0.03 | 0.01 | 0.03 |

# APPENDIX H

## STAGE 4 PASSAGE PACKET ACTIVITY

***These are the solutions: The expected "good" passages are italicized. The principles are underlined.***

**Research Goal:** Identify passages that will aid materials designers in identifying structure-property relationships of biological materials, which can assist in the development of new materials inspired by nature.

**Issue:** Only about 45% of the returned passages are "correctly retrieved", meaning they describe a material structure that may be useful. There are currently too many irrelevant passages, and I want to remove them.

**Instructions:** I need you to look through the following passages and mark each of them as "good" or "bad" and note what property (or properties) the structure would affect. We're looking for structures that may affect strength, ductility, stiffness, toughness, hardness, or density. At the end, please note what you were using to rate the passage and if you notice any patterns that commonly appear in the "good" passages or those that commonly appear in the "bad" passages.

**Here are very clear-cut examples of a good and bad passage. The ones you'll evaluate may be more ambiguous.**

Good Example:

> "…while tightly packed domains tend to resist unfolding, AFM-based single molecule pulling experiments on the muscle protein titin revealed that unraveling of loosely packed modules led to increased molecular contour length and an increase in the energy absorbed prior to catastrophic failure. Similar sacrificial connective proteins have been proposed as toughening agents in the macroscopic mechanical performance of composite biological materials such as nacre and bone…"

TOUGHNESS

Bad Example:

> "The diffraction peaks were indexed according to the inorganic crystal structure database icsd fiz karlsruhe gmbh 2008. Hardness and elastic modulus profiles were obtained by instrumented indentation following the oliver and pharr 2004 method. The applied loads ranged from 014 to 300 mn in twelve successive loading/unloading cycles at increasing loads using a berkovich diamond indenter."

1. *In typical euler buckling formula for columns failure load is directly related to imin such that where is the radius of gyration. Hence by maximizing imin a <u>triangular column</u> makes use of material in a better fashion to yield stiffer structures. The argument is also valid for overall bending rigidity leading to more persistent fibers.*

2. Hydrogels have attracted interest as biomimetic materials due to their pliability extent of hydration low toxicity and biocompatibility. For instance aizenberg and coworkers used hydrogels to mimic the microlens arrays of brittle stars ophiocoma wendii. Hydrogels can also be used to study creatures that creep crawl inch and slither.

3. The results for single level pillars with the shape of punches with rounded edges and an aspect ratio of 1 have been added scale on right ordinate. In this equation f represents the pillar packing density r is the pillar radius and e1n2 is the reduced young's modulus with a poisson ratio of n5 and s is the interfacial strength. Taking s mpa e43 mpa g 068 j m2 and f2267 from reported studies with the same material we obtain geff464 jm2 and 1653 jm2 for l5 and 2.

4. *Some important differences between the reaction wood and normal wood were found for both investigated wood species spruce and yew including the formation of cracks before loading ascribed to residual stresses and the change of fracture mode during crack propagation in the reaction wood. The higher crack propagation resistance was attributed mainly to the <u>different cell ie fiber geometries shape cell wall thickness and fiber angle to the load axis</u> of the reaction wood as basic structural features are responsible for more pronounced crack deflection and branching thus leading to crack growth retardation. Fiber bridging was recognized as another crack growth retarding mechanism which is effective in both wood species and especially pronounced in yew wood.*

5. *This is in stark contrast with monolithic alumina which has a flat curve with KIC 1 MPam1 2. One unique aspect of the architecture of this bioinspired material is that when it has the abalone <u>brick and mortar structure</u> with bridges between the tiles the toughness is increased over the purely lamellar structure 15 MPam1 2. The more conventional simple laminate structure is definitely less tough than the new material.*

6. *The bone plates have a circulatory blood system that attaches to the animal through veins. This unusual structure is similar to lasttiles used on tactical vehicles fig 2b which are <u>hard ceramic hexagonal blocks covered by a polymer</u>. This indicates that the armadillo shell should have excellent impact resistant properties.*

7. *As reported earlier Keunecke et al 2007 2008a Keunecke and Niemz 2008 Keunecke et al 2009 the microfibril angle of yew is rather high and that of reaction wood is even higher. Therefore loading of yew compression wood in the TR direction may be more highly influenced by the microfibril angle than expected so that this microstructural feature may also contribute to the higher fracture tolerance of yew and even more of yew compression wood. In order to verify this assumption more experiments are necessary.*

8. Stress relaxation in a monodisperse system by reptation many biological molecules are long relative to their diameters. A large length to diameter ratio is characteristic of structural molecules in connective tissues and many engineering polymers mechanical properties in a monodisperse system or one where all polymers are assumed to have the same molecular weight are governed by molecules that are long flexible and continuously changing conformation reptation describes this constant motion as very chain at a given instant is confined within a tube as it cannot intersect the neighboring chains. The chain thus moves inside the tube like a snake.

9. *The hydrated ram horn has twice the impact resistance of the horn in the ambient condition. While tensile strength is higher in the ambient state once the horn has been hydrated the strain to failure increases greatly due to the plasticizing effect of water on the amorphous keratin matrix the increased toughness helps the hydrated material absorb more energy before failing. The steer horn behaved like a composite laminate where a cone of damage caused by the impacts was visible.*

10. For the micro cantilevers an additional step was to cut one free end of the beam to achieve a cantilever structure. The width and depth of the cross section of the cantilevers were then measured up to a resolution of 01 musing the sem mode for calculation of the flexural strength and elastic modulus in the micro cantilever bending experiments described below. Some micro cantilevers and beams were made entirely within the enamel or the dentine phase at least 10 m from the dej

11. *The very thin protein layer below serves as a protective film and it can be seen that the next hydrated silica layer is much smoother. Thus what appears to be happening is that the outer thin surfaces can be sacrificial and even if they are tough serve to protect the inner layers and maintain high fracture toughness per a Griffith criterion that correlate with very small crack sizes. Referring back to Eqs 1 and 2 it is important to consider how failure concepts beyond Griffith's equation would apply to ceramic and glass composites such as mollusk shells and sponge spicules.*

12. It is distributed throughout the crown and root and so forms the bulk of tooth and has the function of absorbing and distributing stresses within the tooth.

The structural and compositional dissimilarities between these two mineralised tissues induce marked differences in their mechanical behaviour enamel is much stiffer youngs modulus e 80 gpa and harder hardness h 4 gpa than dentine e 20 gpa h 1 gpa angker and swain 2006. Sharp interfaces between such dissimilar materials are usually subjected to concentrated stresses which often cause delamination.

13. *The difference between mode ii and iii is the shear direction which in this case can be in the normal or transverse direction of the cuticle. A fracture of mode ii has the highest fracture energy because the <u>planes of mineralized fibers are displaced against each other longitudinally</u> with respect to cuticle geometry and the fibers have to be fractured fig 8. In mode i the fracture leads to delaminating of the superimposed mineralized fiber planes and lateral movement.*

14. *Novel unitized structures with porous ti6al4v alloy on one side and compositionally graded hard cocrmo alloy surface on the other side have been fabricated using laser engineered net shaping lens process. <u>Gradient structures</u> with 50% 70% and 86% cocrmo alloy on the top surface showed a high hardness in the range of 615 and 957 hv. The gradient structures were evaluated for their in vitro wear rate and co release up to 3000 m of sliding distance.*

15. *Water molecules are known to form hydrogen bonded bridges within and between the peptide chains and a highly ordered inner hydration layer of water molecules forms hydrogen bonds along the peptide chains to maintain the spacing within the collagen fibrils. The <u>loss of water from collagen fibrils has been postulated to cause the fibrils to become more tightly packed</u> and in turn increase their rigidity due to the larger interfibril attractive forces. Such changes in the structure of collagen have been demonstrated to cause failure of the collagen cross bridging mechanism effective for toughening dentine and to induce unstable brittle failure in the odontogenic tissues.*

16. Due to a well developed pore canal system a honeycomb like structure is generated as numerous canals penetrate the cuticle perpendicular to its surface. Each pore canal contains a long soft and probably flexible tube which has an elliptical like cross section with the long axis of the ellipse parallel to the fiber orientation in each plane fig 3a and b. Additionally the pore canals contain chitin protein fibers oriented perpendicular to those forming the twisted plywood structure.

17. The bending rigidity of a filament is linked to geometrical features of its cross section such as the second moment of area i and the cross sectional area a. For a filament of given cross sectional area a and length l the ability to resist bending due to thermal or mechanical forces is directly proportional to ei where e denotes elastic modulus of the material obtained by measuring the

relationship between force and elongation in a tensile stretching test. For a given filament length and material properties ie elastic modulus e the radius of gyration defined as rg provides a measure of a cross section's capacity to resist bending where imin denotes the smallest second moment of area along any axis passing through the centroid of the cross section.

18. The strain rate sensitivity results showed different behaviours at different strain rates. Based on those behaviours it was found that strain rate related changes in the levels of elastic modulus can be neglected at strain rates higher than 1 s1. This assumption is consistent with elastic modulus values of 103221 gpa obtained at strain rates of 0005 150 s1 wood 1971.

19. *Compared to poisson ratios obtained in the tensile tests which amounted to about 033 in the dry and in the wet state the values for observation direction i are slightly lower. It is remarkable that when compressed in transverse direction the elastic deformation in the normal direction of the cuticle is much smaller than in the transverse direction. This behavior can be explained in terms of the <u>honeycomb like structure</u> of the pore canal system and the direction of the applied stress.*

20. Thus naf changed the apparent mechanical properties of trabecular bone making it less stiff more ductile and less brittle. These changes resulted in a slightly lower average modulus of toughness 031n mm2 in the naf treated group compared to 036n mm2 calculated for the control group. However this reduction was rather small compared to the other detected changes and statistically insignificant.

21. In griffiths theory the elastic energy associated with an applied stress is dissipated into the creation of two new surfaces when the initial crack length is of a critical size c. For the case of crystalline materials capable of plastic deformation the surface energy term in the griffith equation was modified by orowan 1933 to where p is a plastic term and is orders of magnitude larger than in typical natural rigid composite materials comprising primarily ceramics or glass even with a minor matrix component of proteins and other organic matter the plastic term has not been regarded to be of significance as it is in ductile metallic materials. Glass was one of the few materials for which the surface energy is the dominant energy term.

22. *Before complete failure the shell developed hairline cracks that were difficult to see with the unaided eye. This can be attributed to the microstructure which consists of <u>layers of carefully arranged brickwork</u> of calcium carbonate the abalone shell showed brittle fracture and had the second lowest impact strength of 12 kjm2 showing that abalone shell is not naturally optimized for impact. The*

*elk antler exhibited the highest impact strength 58 kjm2 among the mineralized materials.*

23. *After naf treatment another 100 indents were collected on both samples as previously performed. The <u>naf treatment</u> led to significant changes in indentation distance increase idi and elastic modulus in cortical bone. The idi is defined as the increase in the indentation distance in the last cycle relative to the indentation distance in the first cycle of the set of 20 cycles performed.*

24. As a crystal grows it exerts a pressure on its surroundings which depends on the supersaturation of the solution. The gel network can respond to this pressure by rearranging this ability to rearrange is related to the modulus of the gel network and the rate of crystal growth. If the gel network is strong enough to resist the crystallization pressure the crystal will be forced to grow around the gel fibers and hence incorporate them.

25. It is also of interest to note that gordon 1980 suggested that fracture mechanics is not entirely about such matters as work of fracture it is about the communication of energy between the parts or elements of a structure for if the released strain energy cannot reach the fracture zone then the material or the structure cannot break. Two of gordons early major concepts are summed up in fig 12 that illustrates the effect of strong and weak interfaces on crack propagation in a composite material simple examples are also given in gordon 2006. Physically it is of great importance in brittle matrix composites to involve as much of the volume of the specimen under load as possible in the energy dissipating processes.

26. In addition strongly deformed trabeculae exhibit the stress whitening effect which we previously quantified and correlated to microdamage. This is also demonstrated in figre 1b where a region of interest roi of the high speed photograph acquired at 9% apparent strain close to the failure strain of 84% of this sample is compared to scanning electron microscopy sem images of the same roi. In the sem images microcracks with crack bridging were detected.

27. Synthetic materials the 4 ply plywood had impact strength comparable to that of ram horn. The wood dimpled easily and only started to crack and delaminate at impact energies higher than the brittle biological materials bovine femur abalone shell and armadillo carapace. The impactor penetrated through the first two and sometimes third layer before any damage such as cracks or delamination manifested in the bottom layer.

28. Used espi to map displacement patterns of human premolar crowns when loaded on the buccal cusp. The results showed that the enamel cap basically behaves as a stiff body that is capable of deforming and rotating only under

relatively high loads. Here we use espi to map three dimensional displacements deformation rotation of the buccal surface of in situ in the jaw and isolated whole mandibular molars m1 of minipigs as they are being loaded.

29. *The scanning electron micrographs c e illustrate the alignment of both the fibre bundles and the parenchyma with the longitudinal axis of the culm and show that toward the perimphery of the wall the bundles are increasingly closely packed. At higher magnification the fibres are shown to be almost fully dense with a <u>plywood like structure</u> which is due to the internal secondary growth of the cells which increases both density and modulus of the fibres. At this level of structure both the cell walls of the parenchyma and those of the sclerenchyma are polylamellate fibre composites with cellulose fibrils embedded in a hemicellulose and lignin matrix in each layer.*

30. The data on different micropattern sizes are pooled together and presented using box and whisker diagrams as explained in figre 3. Experimental specimens preparations and conditions the biomimetic surface micropattern consisted of hexagonal contact plates of 10 20 50 or 95 mm in size and 1mm in height which were distributed uniformly having the area density of 75% fig 2. Patterned surfaces were produced by pouring twocompound polymerizing pvs coltene whaledent ag altstatten switzerland into the negative templates made of steel ovd kinegram zug Switzerland.

31. *Each value was the mean of five measurements with an error bar indicating the standard deviation. It should be noted that our estimate of <u>porosity</u> is based on the theoretical density of pure cdha 279 gcm3 whereas it could be slightly different when unreacted alpha tcp is present. Nevertheless the latter has a theoretical density of 286 gcm3 therefore the observed amounts of unreacted alpha tcp may not affect the calculated porosity by more than 1%.*

32. *Lastly energetic costs associated with biomineralization in particular of the hypermineralized tooth enameloid layer exist. There are two main strategies for effective defense mechanisms of armor systems defeating threats by means of deforming or fracturing the threat and hence eliminating its ability to penetrate or alternatively providing non-catastrophic sacrificial avenues for deformation and fracture events within the armors microstructure to dissipate the energy of the attack without allowing full penetration. The <u>multilayered microstructural design</u> of the armored scale provides for both of these strategies.*

33. Even in most engineering ceramics is insufficient to predict failure. The attention that has been devoted to studies of the mechanical behavior of rigid natural composites in general in recent years has been due to interest in the combined stiffness strength and toughness of such materials. In particular efforts have addressed mollusk shells and sponge spicules eg vincent and currey

1980 levi et al 1989 sarikaya et al 1995 aizenberg et al 2005 mayer et al 2005a mayer 2005b woesz et al 2006 and a number of other works

34. Using reasonable initial data see termonia 2000 the model predicted at small strains under 0002 a linear curve with modulus e 10 gpa in agreement with experimental observations gosline et al 1994 work 1977 elices et al 2005. At about 002 strain the hydrogen bonds between chains in the amorphous region started to break leading to the formation of a yield point at which the stress strain curve reaches a plateau. At 01 strain and higher the hydrogen bond breaking process was almost complete with the stress resuming its increase with strain.

35. *The rachis consists of a thin cortex 249 12 micron which encloses a relatively thick cellular core measuring from 600 microns to a few millimeters in diameter similar to a <u>sandwich structured composite</u> Fig 2a. This design satisfies resistance to flexure and rupture during flexure without a proportionate increase in weight. The SEM images of dorsal and ventral surfaces reveal relatively smooth topography compared to that of lateral surfaces which exhibit intersecting ridges with a spacing at 10 microns 20 microns with considerable overlap with diameter of cells of the medullary core ranging from 20 microns to 30 microns.*

36. *The presence of a porous cellular architecture is a well documented feature in a variety of natural materials such as wood trabecular bone and plant stems. Their common design of a <u>porous core supporting a denser periphery</u> results in a high specific bending stiffness bending stiffness normalized to material density. However the pore volume fraction is typically much higher 800% than in the sucker rings which are about 20%.*

37. *A dropweight test tower based on standards for testing fiber reinforced polymer matrix composites was designed and fabricated to accommodate the small size of biological materials. The materials tested were divided into two groups <u>nonmineralized</u> and <u>mineralized</u>. The former demonstrated the highest impact strength and showed strong dependence on water content while the latter were relatively brittle and demonstrated no dependence on water content.*

38. *The model is illustrated in figre 1a where stress strain predictions for a range of different <u>fractions of order</u> in a dry generic silk are plotted to show the range of different mechanical properties that are possible for silks. These properties range from a relatively stiff and strong dragline silk with moderate strain to failure all the way to capture silks that have reduced strength but very high toughness through huge strains to failure. Figre 1b shows experimental stress strain curves for a number of different materials for reference.*

39. His end product is artwork but he too must have knowledge of materials science to get there. Chihuly's inspiration and explanations of his work relate glass to water and the shapes of his seaforms to ridged shells with added strength due to shape factor. The names of the colours of glass used in chihuly's work also include inspiration from nature.

40. Calcium radioisotope movement studies on the oyster crassostrea virginica show that movement of the 45ca out of the mantle correlated with the amount of 45ca deposited on the shell growth front. Additional mollusk ion transport studies on the isolated mantle indicate ion movements from the mantle to the shell while other studies suggest that ca2 transport occurs by diffusion through this mantle however this process is not fully understood and studies of this soft tissue can give insights into this biomineralization process. This study intends to investigate the process of mineralization following periods of growth interruption taking into consideration important environmental factors access to food and temperature and to employ high magnification characterization techniques to better understand how the soft tissue (eg epithelium and organic membrane) influences the mechanism of growth.

41. *These experiments have been performed by biological organisms but we have neither investigated many examples nor as a consequence realised their significance. <u>Holes</u> can save money since you need no material to make them make an object lighter make an object more durable control the way a material fails and so make it safer generate information about an object be used to repair an object and by subdividing the structure prepare it for multifunctionality. However there are places where holes are relatively safe.*

42. This indicates that the armadillo shell should have excellent impact resistant properties. Besides having exceptional strength and stiffness given their low density these biological composites must be highly impact resistant. The abalone shell defends the animal against predators who attempt to break their shell or from wave action that causes them to be heaved onto rocks.

43. In order to achieve good accuracy 8 node quadratic plane strain quadrilateral elements were used for the entire mesh. Using interaction energy release rate mode i and mode ii stress intensity factors and consequently the mode mixity were computed. In order to compute accurate stresses at the crack tips collapsed 8 node quadratic quadrilateral elements were used fig 3.

44. Flexure tests show values of 170 mpa compared with a compressive strength of 235 mpa in the same orientation. Dog bone shaped samples were also used to determine the tensile strength of nacre when loaded parallel to the plane of growth the mean strength was considerably lower 65 mpa. Nevertheless the

tensile strength is a much higher fraction of the compressive strength 1/4 1/2 than in monolithic ceramics where it fluctuates between 1/10 and 1/15.

45. This effect had previously only been demonstrated for the incorporation of individual particles and was not appreciated for gel networks. By controlling both gel strength and crystal growth kinetics we have achieved complete incorporation and total exclusion of the gel network for a single gel crystal pair calcite and agarose. These results have the potential to lead to design criteria for polymer reinforced crystalline materials with unique structure property relationships.

46. In addition lanthanides are relevant additives for the uhmwpe because they have been suggested as tracers for polyethylene wear kunze and wimmer 2006 ngai et al 2009 wimmer et al 2006. A recent study on the oxidation products in the uhmwpe doped with the euiii stearate found decreased carbonyl and hydroperoxide generation rates gallardo et al in press laurent et al 2010 with oxidation retardation effects akin to those reported for the photodegradation of a high density polyethylene doped with copper stearate osawa et al 1979. Lanthanides are also used as a replacement for antibiotics in certain agricultural sectors of china and switzerland redling 2006.

47. *This is most likely related to differences in the structure <u>mineral content</u> and morphology of the respective enamels. Previous studies comparing human and pig molar crowns have shown that despite overall similarities in morphology development and function human teeth have a different enamel microstructure as well as a stiffer thicker and less complex enamel cap compared to the pig. The enamel of pig teeth is only 80% mineralized at maturity compared to 97% mineralization in human enamel.*

48. *The palms Wlfia and Iriartea which have a density <u>gradient</u> across the section with n 5 and n 10 respectively and typical values 1 2 3 and 1 31 1 38 have stems that are 1 5 2 times lighter or more efficient than a solid circular beam made from solid cell wall material. Bamboos with a quadratic density distribution across the section and typical values 2 6 2 and 1 17 1 05 respectively have culms that are 2 4 times more efficient than a solid circular beam made from solid cell wall material. 4. Materials with property gradients in engineering design Tradition and current high volume use particularly of bamboo is in construction both as a building material and a scaffolding material.*

49. The yield strength is usually determined by its linear relationship with the hardness h 28y fischer cripps 2004 so it can be easily calculated for the ti substrate. Since the layer hardness is the concern of this study a different approach is needed for the layer yield strength yf soares et al 2008 presented a methodology for yf calculation as follows. The titania layer can be considered as

a ceramic layer and its yf value can be obtained by the relationship presented by milman and chugunova 1999 where 49 and 21 are numeric parameters and h denotes the materials plasticity.

**Please note what you were using to evaluate the passage.**
**Also, what patterns did you notice commonly appear in the “bad” passages as opposed to the “good” passages?**

## APPENDIX I

## STAGE 4 ALGORITHM CODE

### BioinspiredMaterialsMain Class

```
import java.io.File;
import java.io.FileWriter;
import java.io.IOException;
import java.io.PrintWriter;
import java.util.ArrayList;
import java.util.Arrays;
import java.util.Scanner;

public class BioinspiredMaterialsMain {

	private static final String[] strengthPhrases = {"load carrying capacities", "load carrying
capacity", "modulus of rupture", "strength", "strengthen",
			"strengthener", "strengthening", "strengths", "stronger", "tensile force", "tensile forces",
"yield point", "yield points", "maximum stress",
			"maximum stresses", "yield stress", "yield stresses"};
	private static final String[] ductilityPhrases = {"brittle", "brittlely", "brittleness", "brittler",
"brittlest", "ductile", "ductilely", "ductilities",
			"ductility", "extensibilities", "extensibility", "extensible", "failure length", "failure
lengths", "failure strain", "failure strains", "fracture length",
			"fracture lengths", "fracture strain", "fracture strains", "original gauge length", "original
gauge lengths", "percent reduction in area",
			"percent reduction in areas", "percent reductions in area", "strain to failure", "strain to
fracture", "strains to failure", "strains to fracture"};
	private static final String[] stiffnessPhrases = {"elastic deformation", "elastic deformations",
"hookes law", "Hooke's law", "moduli", "modulus",
			"resistance to bending", "resistances to bending", "resistant to bending", "rigid",
"rigidities", "rigidity", "rigidly", "rigidness", "stiff",
			"stiffen", "stiffened", "stiffening", "stiffer", "stiffest", "stiffly", "stiffness", "stiffnesses",
"strain energy"};
	private static final String[] toughnessPhrases = {"absorb energy", "absorbing energies",
"absorbing energy", "absorbs energy", "absorption of energy",
			"crack propagation", "crack propagations", "dissipate energies", "dissipate energy",
"dissipated energy", "dissipates energy", "dissipation of energy",
			"energies absorbed", "energy absorbed", "energy absorption", "energy dissipated",
"energy dissipation", "fracture energies", "fracture energy",
			"fracture resistance", "fracture resistances", "fracture resistant", "impact resistance",
"impact resistances", "impact resistant", "impact strength",
			"impact strengths", "resist fracture", "resistance to fracture", "resistances to fracture",
"resistant to fracture", "resists fracture", "strain energies",
			"strain energy", "tough", "toughen", "toughened", "toughening", "tougher", "toughest",
"toughness", "toughnesses"};
	private static final String[] hardnessPhrases = {"hard", "harden", "hardening", "harder",
"hardest", "hardness", "hardnesses", "indentation map",
			"indentation maps", "microindentation map", "microindentation maps",
"nanoindentation map", "nanoindentation maps", "soft", "soften", "softening",
```

```
"softer", "softest", "softness", "softnesses"};
	private static final String[] densityPhrases = {"buoyancy", "dense", "densest", "densities",
"density", "lightening", "specific bending stiffness",
		"specific volume", "specific volumes", "weight", "weights"};

	private static final String[] strengthStandAlone = {"strengthen", "strengthened",
"strengthening", "stronger", "strongest", "weakened", "weakens", "weaker", "weakest"};
	private static final String[] ductilityStandAlone = {};
	private static final String[] stiffnessStandAlone = {"resistant to bending", "stiffen", "stiffening",
"stiffer", "stiffest"};
	private static final String[] toughnessStandAlone = {"crack propagation", "toughen",
"toughened", "toughening", "toughens", "tougher", "toughest", "toughness"};
	private static final String[] hardnessStandAlone = {"harden", "hardened", "hardening",
"hardens", "harder", "hardest", "soften", "softened", "softening", "softens", "softer", "softest"};
	private static final String[] densityStandAlone = {"denser", "densest", "densified", "lightening",
"lightweight"};

	private static final String[] strengthExclude = {"absorption strength", "actuation stress",
"adhesion strength", "adhesion strengths",
		"adhesive strength", "adhesive strengths", "after the yield point", "applied stress",
"applied stresses", "before the yield point",
		"beyond the yield point", "binding strength", "binding strengths", "bond strength", "bond
strengths", "bonding strength",
		"bonding strengths", "chiral strength", "clamping stress", "compatibility stresses",
"constant stress", "contact strength",
		"coupling strength", "coupling strengths", "creep strengthening", "crystal strength",
"cyclic stress", "cyclic stresses", "dielectric strength",
		"electrostatic stress", "epitaxial stress", "exchange strength", "fatigue strength", "fatigue
strengths", "fatigue stress", "fatigue stresses",
		"field strength", "field strengths", "filter strength", "flow stress", "frictional strength",
"full strength", "gel strength", "gel strengths",
		"generated stress", "hybridization strength", "impact strength", "impact strengths",
"important to stress", "initiation stress",
		"interaction strength", "interaction strengths", "interfacial strength", "interfacial
strengths", "interfacial stress", "interfacial stresses",
		"intergranular stress", "intergranular stresses", "interlaminar strength", "internal stress",
"internal stresses", "ionic strength",
		"ionic strengths", "key strength", "major strength of", "nucleation stress", "optical
strength", "oscillator strength", "oscillator strengths",
		"osmotic stress", "peierls stress", "periodic stress", "pull off strength", "pull off
strengths", "readings of stress", "relaxation of stresses",
		"relaxation strength", "relaxation strengths", "relaxes stresses", "residual stress",
"residual stresses", "shear stress tau",
		"signal strengths", "strength of adhesion", "strength of chirality", "strength of
interaction", "strength of magnetic",
		"strength of magnetization", "strength of superconductivity", "strength of the", "strength
of these", "strength tester",
		"strengths and weaknesses", "strengths in", "stress amplitude", "stress corrosion", "stress
free", "stress function", "stress generation",
		"stress intensities", "stress intensity", "stress plot", "stress range", "stress rate", "stress
reading", "stress recovery", "stress relaxation",
		"stress relaxes", "stress sigma", "stress strain", "stress tensors", "stress that", "stress the
importance", "stress whitening", "stresses of",
```

```
                    "stresses the importance", "stronger approach", "stronger attraction", "stronger
attractive", "stronger binding", "stronger bonding",
                    "stronger dependence", "stronger dipoles", "stronger impact", "stronger increase",
"stronger influence", "stronger interaction",
                    "stronger interfacial", "stronger tendency", "stronger volume", "stronger wear", "surface
stress", "surface stresses", "the stronger the",
                    "to stress", "transition strength", "transition strengths", "unit of stress", "up to a yield
point", "viscous stresses", "which stresses how",
                    "zero stress"};
          private static final String[] ductilityExclude = {"brittle star", "brittle stars", "fatigue ductility"};
          private static final String[] stiffnessExclude = {"assume elastic deformation", "bending modulus
eb", "bulk modulus b", "bulk modulus k",
                    "by means of hookes law", "chain rigidity", "complex moduli", "complex modulus",
"contact moduli", "contact modulus", "contact stiffness",
                    "dynamic moduli", "dynamic modulus", "elastic deformation energy", "elastic
deformation fields", "elastic modulus e", "elimination modulus",
                    "equilibrium elastic deformation", "exchange stiffness", "fourier moduli", "fourier
modulus", "gaussian bending rigidity", "indentation moduli",
                    "indentation modulus", "loss moduli", "loss modulus", "magnetic stiffness", "moduli
data", "modulus g", "modulus measurements",
                    "modulus of capacitance", "modulus of rupture", "modulus of toughness", "molecular
rigidity", "plateau moduli", "plateau modulus",
                    "purely elastic deformations", "range of moduli", "repeated elastic deformation", "rigid
backbone", "rigid body", "rigid building", "rigid chain", "rigid groups",
                    "rigid molecular", "rigid molecules", "rigid rod", "rigid rotation", "rigid rotations", "rigid
shift", "rigidly fixed", "rigidly shearing",
                    "rubber modulus", "rubbery moduli", "rubbery modulus", "shear modulus g", "stiff
indenter", "storage moduli", "storage modulus",
                    "stretch moduli", "stretch modulus", "tilt moduli", "tilt modulus", "tilt stiffness",
"viscous moduli", "viscous modulus", "weibull modulus",
                    "youngs modulus e", "youngs modulus y"};
          private static final String[] toughnessExclude = {"crack propagation rate", "dissipated energy
density", "dissipation of energy flux",
                    "energy absorption band", "energy absorption peak", "energy dissipation density",
"indentation toughness", "interface fracture energies",
                    "interface fracture energy", "interface fracture resistance", "interface fracture
resistances", "interface fracture toughness",
                    "interface toughness", "interfacial fracture energies", "interfacial fracture energy",
"interfacial fracture resistance",
                    "interfacial fracture resistances", "interfacial fracture toughness", "interfacial
toughness", "interlaminar fracture toughness",
                    "ion energy dissipation", "measurement of toughness", "molecular energy dissipation",
"rate of crack propagation", "strain energy function",
                    "tackle tough", "tough competition", "tough decision", "tough decisions", "tough game",
"tough one", "tough problem", "tough problems",
                    "tough to", "toughening agent", "toughening agents", "tougher criteria", "toughness gic",
"toughness giic", "toughness kc", "toughness kic",
                    "toughness measurement", "toughness measurements", "toughness parameter",
"toughness testing", "toughness value", "toughness values",
                    "velocity toughening"};
          private static final String[] hardnessExclude = {"beam hardening", "cyclic hardening", "hard
axes", "hard axis", "hard choice", "hard choices",
```

```
			"hard contact", "hard contacts", "hard core", "hard disk", "hard disks", "hard drive",
"hard drives", "hard indenter", "hard light",
			"hard matter", "hard question", "hard questions", "hard sphere", "hard spheres", "hard
tapping", "hard template", "hard templates",
			"hard templating", "hard thing", "hard tip", "hard to", "hard way", "hard wearing", "hard
x ray", "hard x rays", "harder contact",
			"harder contacts", "harder for", "harder than", "harder to", "hardness h", "hardness hk",
"hardness hv", "hardness indentation",
			"hardness indentations", "hardness indenter", "hardness indent", "hardness indents",
"hardness scale", "hardness tester", "indentation hardness",
			"knoop hardness", "magnetically soft", "microhardness", "precipitation hardening",
"radiation hardening", "soft adhesion", "soft bake",
			"soft baked", "soft chemistry", "soft cloth", "soft contact", "soft contacts", "soft
elasticity", "soft eye", "soft imaging", "soft knee",
			"soft landed", "soft landing", "soft lithographic", "soft lithography", "soft magnetic",
"soft magnets", "soft matter", "soft method",
			"soft mode", "soft molding", "soft photolithography", "soft science", "soft sphere", "soft
system", "soft systems", "soft tapping",
			"soft technique", "soft template", "soft templates", "soft templating", "soft x ray", "soft x
rays", "softening temperature",
			"solid solution hardening", "strain hardening", "strain rate hardening", "strain softening",
"stress hardening", "vickers hardness",
			"water hardness", "work hardening", "work softening"};
	private static final String[] densityExclude = {"absorption density", "anion densities", "areal
densities", "areal density", "atom density",
			"atomic densities", "atomic density", "atomic number densities", "atomic weight",
"atomic weights", "band densities", "band density",
			"bit densities", "bit density", "branch densities", "branch density", "by weight", "calorie
dense", "capacitance densities",
			"capacitance density", "carrier densities", "carrier density", "cation density", "cell
densities", "cell density", "chain densities",
			"chain density", "charge densities", "charge density", "cluster densities", "cluster
density", "crack densities", "crack density",
			"crosslink densities", "crosslink density", "crosslinking densities", "crosslinking
density", "crystallographic densities",
			"current densities", "current density", "data density", "defect dense", "defect densities",
"defect density", "dense arms", "dense array",
			"dense arrays", "dense bend contours", "dense cell", "dense concentration", "dense
crystal packing", "dense defects", "dense DNA packing",
			"dense fibrillar matrix", "dense flow", "dense flows", "dense framework", "dense
frameworks", "dense grains", "dense granular flows",
			"dense group", "dense groups", "dense inclined flows", "dense integration", "dense
lattice", "dense limit", "dense mesh", "dense meshes",
			"dense molecular packing", "dense nanotube arrays", "dense network", "dense
nucleation", "dense packing", "dense packings",
			"dense parallel arrays", "dense pattern", "dense patterns", "dense periodic", "dense
plane", "dense planes", "dense population",
			"dense populations", "dense random", "dense regular packing", "dense spots", "denser
arrays", "denser concentration", "denser grid",
			"denser nucleation", "denser packing", "denser plane", "densest arrays", "densest
concentration", "densest packing", "densest plane",
			"densities of cells", "densities of charge", "densities of impurities", "densities of state",
"densities of states", "density array",
```

```
			"density arrays", "density carriers", "density dislocations", "density driven", "density functional", "density of carriers", "density of charge",
			"density of clusters", "density of crosslinks", "density of defects", "density of dipoles", "density of dislocations", "density of hairs",
			"density of particles", "density of point defects", "density of pores", "density of power", "density of state", "density of states",
			"density of surface defects", "density of the dislocations", "density packing", "density relaxation", "density storage", "density waves",
			"deposition densities", "dipole density", "dislocation densities", "dislocation density", "DNA densities", "dopant densities", "dopant density",
			"doping densities", "dot densities", "drop weight", "dry weight", "electrode density", "electron dense", "electron densities", "electron density",
			"electronic densities", "electronic density", "electronically denser", "energy dense", "energy densities", "energy density", "entanglement densities",
			"entanglement density", "excitation density", "fault densities", "fault density", "flux densities", "flux density", "framework densities",
			"framework density", "graft densities", "graft density", "grafting densities", "grid densities", "grid density", "hairpin densities",
			"hairpin density", "hole densities", "hole density", "image density", "information density", "integration densities", "interlink densities",
			"interlink density", "junction densities", "junction density", "length densities", "ligand densities", "ligand density", "local density approximation",
			"magnetization densities", "magnetization density", "memory densities", "memory density", "mineral densities", "mineral density", "molecular densities",
			"molecular density", "molecular weight", "molecular weights", "momentum density", "nucleation densities", "nucleation density", "nucleus density",
			"number density", "optical densities", "optical density", "ordering densities", "packaging densities", "packed densities", "packed density",
			"packing densities", "packing density", "pair densities", "particle densities", "particle density", "particle number densities",
			"photocurrent densities", "photocurrent density", "pinning densities", "pinning density", "probability weight", "polarization densities",
			"polarization density", "polymer weight ratio", "polymer weight ratios", "population densities", "population density", "pore densities",
			"pore density", "power dense", "power densities", "power density", "precipitate densities", "precipitate density", "probability densities",
			"probability density", "receptor densities", "receptor density", "recording densities", "recording density", "seeding densities", "seeding density",
			"segment densities", "segment density", "site density", "spatial densities", "spatial density", "spectral densities", "spectral density",
			"spectral weight", "spin densities", "spin density", "stack densities", "stack density", "stacking densities", "stacking density", "storage densities",
			"storage density", "supercurrent density", "surface densities", "surface density", "thread densities", "thread density", "transistor density",
			"vortex density", "weight average", "weight averaged", "weight bearing", "weight concentration", "weight concentrations", "weight distribution",
			"weight distributions", "weight fraction", "weight fractions", "weight percent", "weight percentage", "weight ratio", "weight ratios",
			"weight sensitivity", "weight sensitivities"};

	public static void main(String[] args) throws IOException
	{
		Scanner kbInput = new Scanner(System.in);
```

```
		long startTime = System.currentTimeMillis();

		{
			System.out.println("!!PLEASE MAKE SURE TO READ THE READ-ME
FILE BEFORE USING THE PROGRAM!!");
			System.out.println("Please enter number(s) of the desired property: [12 for
strength and ductility]");
			System.out.println("1. strength\n2. ductility\n3. stiffness\n4. toughness\n" +
							"5. hardness\n6. density");
			String choices = kbInput.nextLine();

			ArrayList<ArrayList<String>> searchTermLists = selectSearchTerms(choices);
			ArrayList<ArrayList<String>> exclusionTermLists =
selectExclusionTerms(choices);
			ArrayList<ArrayList<String>> standAloneTermLists =
selectStandAloneTerms(choices);

			String fileName = getFilename(choices);
			PrintWriter writer = new PrintWriter(new FileWriter(fileName, false));

			File folder = new File("./journals");
			for(String s : folder.list())
				System.out.println(s);

			writer = new PrintWriter(new FileWriter(fileName, true));
			BioLibrary library;
			for(File file: folder.listFiles())
			{
				library = new BioLibrary(file, searchTermLists, exclusionTermLists,
standAloneTermLists);
				String result = library.searchArticles();
				if(result.equals(""))
				{
					result = "No result found for the property(s) selected in "+
file;
					System.out.println(result);
				}
				else
					System.out.println(result);
				System.out.println("Number of Articles searched in " + file + " is " +
library.getNumberOfArticlesSearched());

				writer.println(file + "\n\n" + result);
			}
			writer.close();
		}
		kbInput.close();
		long endTime = System.currentTimeMillis();
		System.out.println("Search took " + (endTime - startTime) + " milliseconds");
	}

	private static String getFilename(String s)
	{
```

```java
		String temp = "";
		for(int i = 0; i < s.length(); i++)
		{
			char c = s.charAt(i);
			switch(c-48)
			{
				case 1:
					temp += "strength";
					break;
				case 2:
					temp += "ductility";
					break;
				case 3:
					temp += "stiffness";
					break;
				case 4:
					temp += "toughness";
					break;
				case 5:
					temp += "hardness";
					break;
				case 6:
					temp += "density";
					break;
			}
			if(!(i==s.length()-1))
				temp += " ";
		}
		temp += ".txt";
		return temp;
	}

	private static ArrayList<ArrayList<String>> selectSearchTerms(String s)
	{
		ArrayList<ArrayList<String>> temp = new ArrayList<ArrayList<String>>();
		for(char c : s.toCharArray())
		{
			switch(c-48)
			{
				case 1:
					temp.add(new
ArrayList<String>(Arrays.asList(strengthPhrases)));
					System.out.println(c);
					break;
				case 2:
					temp.add(new
ArrayList<String>(Arrays.asList(ductilityPhrases)));
					System.out.println(c);
					break;
				case 3:
					temp.add(new
ArrayList<String>(Arrays.asList(stiffnessPhrases)));
					System.out.println(c);
```

```
                                        break;
                                case 4:
                                        temp.add(new
ArrayList<String>(Arrays.asList(toughnessPhrases)));
                                        System.out.println(c);
                                        break;
                                case 5:
                                        temp.add(new
ArrayList<String>(Arrays.asList(hardnessPhrases)));
                                        System.out.println(c);
                                        break;
                                case 6:
                                        temp.add(new
ArrayList<String>(Arrays.asList(densityPhrases)));
                                        System.out.println(c);
                                        break;
                        }
                }
                return temp;
        }

        private static ArrayList<ArrayList<String>> selectExclusionTerms(String s)
        {
                ArrayList<ArrayList<String>> temp = new ArrayList<ArrayList<String>>();
                for(char c : s.toCharArray())
                {
                        switch(c-48)
                        {
                                case 1:
                                        temp.add(new
ArrayList<String>(Arrays.asList(strengthExclude)));
                                        System.out.println(c);
                                        break;
                                case 2:
                                        temp.add(new
ArrayList<String>(Arrays.asList(ductilityExclude)));
                                        System.out.println(c);
                                        break;
                                case 3:
                                        temp.add(new
ArrayList<String>(Arrays.asList(stiffnessExclude)));
                                        System.out.println(c);
                                        break;
                                case 4:
                                        temp.add(new
ArrayList<String>(Arrays.asList(toughnessExclude)));
                                        System.out.println(c);
                                        break;
                                case 5:
                                        temp.add(new
ArrayList<String>(Arrays.asList(hardnessExclude)));
                                        System.out.println(c);
                                        break;
```

```
				case 6:
					temp.add(new
ArrayList<String>(Arrays.asList(densityExclude)));
					System.out.println(c);
					break;
			}
		}
		return temp;
	}

		private static ArrayList<ArrayList<String>> selectStandAloneTerms(String s)
		{
			ArrayList<ArrayList<String>> temp = new ArrayList<ArrayList<String>>();
			for(char c : s.toCharArray())
			{
				switch(c-48)
				{
					case 1:
						temp.add(new
ArrayList<String>(Arrays.asList(strengthStandAlone)));
						System.out.println(c);
						break;
					case 2:
						temp.add(new
ArrayList<String>(Arrays.asList(ductilityStandAlone)));
						System.out.println(c);
						break;
					case 3:
						temp.add(new
ArrayList<String>(Arrays.asList(stiffnessStandAlone)));
						System.out.println(c);
						break;
					case 4:
						temp.add(new
ArrayList<String>(Arrays.asList(toughnessStandAlone)));
						System.out.println(c);
						break;
					case 5:
						temp.add(new
ArrayList<String>(Arrays.asList(hardnessStandAlone)));
						System.out.println(c);
						break;
					case 6:
						temp.add(new
ArrayList<String>(Arrays.asList(densityStandAlone)));
						System.out.println(c);
						break;
				}
			}
			return temp;
		}
}
```

**BioLibrary Class**

```
import java.io.BufferedReader;
import java.io.File;
import java.io.FileReader;
import java.io.IOException;
import java.util.ArrayList;

//BioLibrary class - takes the searchTermLists from main, opens a text file and searches through the
articles
//responsible for:
//1. parsing through the article for sentences and create Article object
//2. calling the search algorithm from Article

public class BioLibrary {
	private BufferedReader fileInput;
	private Article currentArticle;

	private ArrayList<ArrayList<String>> searchTermLists;
	private ArrayList<ArrayList<String>> exclusionTermLists;
	private ArrayList<ArrayList<String>> standAloneTermLists;
	private String searchResult;

	private boolean getInfo;
	private boolean getContent;

	private int numArticles;

	public BioLibrary(File file, ArrayList<ArrayList<String>> list1, ArrayList<ArrayList<String>>
list2, ArrayList<ArrayList<String>> list3) throws IOException
	{
		fileInput = new BufferedReader(new FileReader(file));
		searchTermLists = list1;
		exclusionTermLists = list2;
		standAloneTermLists = list3;

		currentArticle = null;
		searchResult = "";
		getInfo = false;
		getContent = false;

		findFirstArticle();
	}

	private void findFirstArticle() throws IOException
	{
		String line = "";

		while((line = fileInput.readLine()) != null)
		{
			if(line.equals("##"))
				break;
		}
```

```
		getInfo = true;
		getContent = false;
	}

	private void loadNext() throws IOException
	{
		if(!getInfo)
		{
			currentArticle = null;
			return;
		}

		String line = "";
		String info = "";

		Sentence tempSentence = new Sentence();

		while((line = fileInput.readLine()) != null)
		{
			if(line.equals("##"))
			{
				getContent = false;
				getInfo = true;
				info = "";
				break;
			}

			if(getInfo)
			{
				if(line.equals("#"))
				{
					getInfo = false;
					getContent = true;
					currentArticle = new Article(info);
				}
				else
				{
					line = line.trim();
					if(!line.equals(""))
					{
						info += line;
						info += "\n";
					}
				}
				continue;
			}

			if(getContent)
			{
				line = line.trim();

				for(String s : line.split("\\s+|-"))
```

```
                        {
                                if(s.equals(""))
                                        continue;

                                s = s.replaceAll("[F|f]ig\\.", "fig");
                                s = s.replaceAll("[F|f]igs\\.", "figs");
                                s = s.replaceAll("[E|e]q\\.", "eq");
                                s = s.replaceAll("[E|e]qs\\.", "eqs");
                                s = s.replaceAll("al\\.", "al");
                                s = s.replaceAll("etc\\.", "etc");
                                s = s.replaceAll("i\\.e\\.", "ie");
                                s = s.replaceAll("e\\.g\\.", "eg");
                                s = s.replaceAll("\\[.+\\]", "");
                                s = s.replaceAll("[^a-zA-Z0-9_.?!%/\'\"]", "");
                                s = s.replaceAll("%", " percent");

                                if(s.equals(""))
                                        continue;
                                char lastChar = s.charAt(s.length()-1);

                                s = s.replaceAll("[.?!]", "");
                                if(s.equals(""))
                                        continue;

                                if(lastChar == '.' || lastChar == '?' || lastChar == '!')
                                {
                                        tempSentence.addWord(s.replaceAll("[.?!]", ""));
                                        if(tempSentence.length() > 2)
                                                currentArticle.addSentence(tempSentence);
                                        tempSentence = new Sentence();
                                }
                                else
                                {
                                        tempSentence.addWord(s);
                                }
                        }
                }
        }
        if(!tempSentence.toString().equals(""))
                currentArticle.addSentence(tempSentence);
}

//runs search algorithm
public String searchArticles() throws IOException
{
        loadNext();
        while(currentArticle != null)
        {
                numArticles++;
                if(currentArticle.isBioRelated())
                {
```

```
					String temp = currentArticle.runSearchAlgorithm(searchTermLists, 
exclusionTermLists, standAloneTermLists);
					if(!temp.equals(""))
					{
						searchResult += "Article Info:\n" + currentArticle.getInfo();
						searchResult += "Search Result:\n" + temp + "\n\n";
					}
				}
				loadNext();
			}
			return searchResult;
		}

		public int getNumberOfArticlesSearched()
		{
			return numArticles;
		}
}
```

## Article Class

```
import java.util.ArrayList;
import java.util.Arrays;
import java.util.HashSet;
import java.util.TreeSet;
import java.util.regex.Matcher;
import java.util.regex.Pattern;

//Article class - saves the biographical info and sentences of an article
//responsible for running searching algorithm

public class Article {

		private static final String[] bioPhrases = {"biological", "biology", "biopolymer", "biopolymers",

		"biomaterial", "biomaterials", "bioscience", "life",

		"living", "natural", "nature", "organic", "organism",

		"organisms", "protein", "proteins", "tissue", "tissues",

		"phylum", "genus", "species", "plant", "plants", "animal",

		"animals"};
		private static final String[] identifierPhrases = {"a measure of", "adding to", "adjusting", 
"alter", "considerable", "decrease",
			"decreased", "decreases", "decreasing", "degree of", "enhanced", "excellent", 
"exceptional", "exceptionally", "exponential", "full",
			"gain", "good", "greater", "high", "higher", "highest", "improve", "improved", 
"improving", "increase", "increased", "increases",
			"increasing", "large", "less", "low", "lower", "lowered", "lowers", "lowest", "maximum", 
"minimizes", "minimum", "more", "most",
```

```
			"negative", "optimized", "outstanding", "per cent", "percent", "reduce", "reduced", "reduces", "reducing", "reduction", "reinforcement",
			"same", "significant", "similar", "smaller", "sufficient", "times as", "twice as", "variation", "varying", "correlated to", "correlation",
			"correlations", "corresponding", "dependence of", "dependence on", "dependent on",
			"govern the", "influence on", "link between", "provides", "providing", "role in", "role of", "roles in", "ruled by"};
	private static final String[] relationshipPhrases = {"ability of", "ability to", "achieved", "achieving", "act as", "acts as", "acts to",
			"affecting", "affects", "allows for", "ascribed to", "associated", "attributable to", "attributed", "because", "by", "cause a",
			"caused", "causing", "connected to", "consequence of", "consequently", "contribute to", "contributes to", "contributing to", "contribution",
			"convey", "correlate",
			"correspond to", "corresponds to", "demonstrated", "depend on", "dependences", "depending on", "depends on", "due to", "effect",
			"effects", "enable", "evidence", "exhibited", "explain", "explained", "explains", "function of", "gave", "give", "gives", "imparts",
			"implying that", "in order to", "inducing", "influence of", "lead to", "leading", "leads to", "led to", "linked to", "makes", "making",
			"means to", "modifies", "produces", "producing", "promotes", "provide", "provoke", "reason", "related", "relation", "responsible for",
			"result in", "result of", "resulted in", "resulting from", "resulting in", "results in", "satisfies", "shown to", "shows how", "since",
			"thereby", "therefore", "thus", "to affect", "to cause", "to control", "to induce", "to make", "to produce", "to promote", "to yield",
			"used to", "which", "why", "with", "yielded", "yielding", "yields", "correlated to", "correlation",
			"correlations", "corresponding", "dependence of", "dependence on", "dependent on", "govern the", "influence on", "link between", "provides",
			"providing", "role in", "role of", "roles in", "ruled by"};
	private static final String[] comparisonPhrases = {"changed from", "compared to", "difference between", "in comparison", "in contrast to",
			"in contrast with", "than", "the other hand", "when comparing", "whereas", "while"};

	private static final String[] procedurePhrases = {"proctest1", "a constant", "crosshead", "cycle", "cycles", "experimental",
			"extensometer", "frequencies", "frequency", "indentation load", "instrumented", "load application", "loading unloading",
			"loadingunloading", "measured", "measurement", "measurements", "mold", "molds", "rate", "rates", "sampled from", "setup",
			"specimen", "specimens", "test", "tested", "testing", "tests"};
	private static final String[] calculationPhrases = {"calctest1", "area under", "calculated", "compute", "defined as", "determined by",
			"determined for", "parameter", "parameters", "taken as", "taken to be", "e is the elastic modulus", "energy g", "energy gf",
			"hardness h", "length l", "load fmax", "mass m", "modulus e", "modulus et", "rigidity ei", "stiffness e", "stiffness ei",
			"strength yf", "stress y", "y the yield strength"};
	private static final String[] resultsPhrases = {"resultstest1", "analytical method", "analytical methods", "assumed", "curve", "curves",
			"data", "slope", "slopes", "values"};
```

```java
	private static final int identifierRange = 0;
	private static final int relationshipRange = 1;
	private static final int comparisonRange = 1;
	private static final int procedureRange = 1;
	private static final int calculationRange = 1;
	private static final int resultsRange = 0;

	private String info;
	private boolean bioRelated;

	private ArrayList<Sentence> sentences;

	public Article(String i)
	{
		info = i;
		bioRelated = false;
		sentences = new ArrayList<Sentence>();
	}

	public void addSentence(Sentence s)
	{
		sentences.add(s);
		if(!bioRelated && s.matchPhrases(new ArrayList<String>(Arrays.asList(bioPhrases))))
		{
			bioRelated = true;
		}
	}

	public boolean isBioRelated()
	{
		return bioRelated;
	}

	public String getInfo()
	{
		return info;
	}

	public String toString()
	{
		String content = "";
		for(Sentence s : sentences)
		{
			content += s.toString();
			content += "\n";
		}
		return "Article Info:\n" + info + "\nContent:\n" + content;
	}

	//run search algorithm and return the resulting sentences
	public String runSearchAlgorithm(ArrayList<ArrayList<String>> searchTermList,
ArrayList<ArrayList<String>> exclusionTermList, ArrayList<ArrayList<String>> standAloneTermList)
	{
```

```
        ArrayList<HashSet<Integer>> matchingIndexes =
getPropertyMatchingIndexes(searchTermList, exclusionTermList);
        ArrayList<HashSet<Integer>> matchingSAIndexes =
getPropertyMatchingIndexes(standAloneTermList, exclusionTermList);

        HashSet<Integer> identifierIndexes = getHelperIndexes(identifierPhrases,
identifierRange, 0);

        HashSet<Integer> relationshipIndexes = getHelperIndexes(relationshipPhrases,
relationshipRange, 1);
        HashSet<Integer> relationshipIfThenIndexes = getIfThenIndexes(relationshipRange,
0);
        relationshipIndexes.addAll(relationshipIfThenIndexes);
        HashSet<Integer> comparisonIndexes = getHelperIndexes(comparisonPhrases,
comparisonRange, 0);

        HashSet<Integer> propIdMergedIndexes =
getMergedIndexes(multiPropertyHash(matchingIndexes), identifierIndexes);
                //that contain a property term AND a modification term (getIdentifierIndexes
method) in the same sentence
                //if multiple properties, the property terms must appear in the same sentence
        HashSet<Integer> propIdRMergedIndex = getMergedIndexes(propIdMergedIndexes,
relationshipIndexes);
                //1. causation + modifier + property
                //2. correlation + property
        HashSet<Integer> propIdCompMergedIndex =
getMergedIndexes(propIdMergedIndexes, comparisonIndexes);
                //3. comparison + modifier + property
        HashSet<Integer> standAloneIndexes = multiSAPropertyHash(matchingSAIndexes,
matchingIndexes);
                //JT for multiple properties (SA and regular terms)
        HashSet<Integer> standAloneRMergedIndex = getMergedIndexes(standAloneIndexes,
relationshipIndexes);
                //4. causation + stand-alone
                //5. correlation + stand-alone
        HashSet<Integer> standAloneCompMergedIndex =
getMergedIndexes(standAloneIndexes, comparisonIndexes);
                //6. comparison + stand-alone

        //Creating a mergedIndexes hashset with all of the sentence indices of interest
        HashSet<Integer> mergedIndexes = new HashSet<Integer> ();
        mergedIndexes.addAll(propIdRMergedIndex);
        mergedIndexes.addAll(propIdCompMergedIndex);
        mergedIndexes.addAll(standAloneRMergedIndex);
        mergedIndexes.addAll(standAloneCompMergedIndex);

        //Removing the indices of the bad pattern terms
        //Range = 0 if the term cannot appear in the property sentence
        //Range = 1 if the term cannot appear in the passage
        HashSet<Integer> procedureIndexes = getHelperIndexes(procedurePhrases,
procedureRange, 0);
```

```
		HashSet<Integer> calculationIndexes = getHelperIndexes(calculationPhrases, 
calculationRange, 0);
		HashSet<Integer> resultsIndexes = getHelperIndexes(resultsPhrases, resultsRange, 0);
		mergedIndexes.removeAll(procedureIndexes);
		mergedIndexes.removeAll(calculationIndexes);
		mergedIndexes.removeAll(resultsIndexes);

		TreeSet<Integer> returnIndexes = new TreeSet<Integer>(mergedIndexes);
		returnIndexes = removeNumbers(returnIndexes);
			//comment out above line if you want to keep the passages that have numbers in 
the property sentence

		String returnSentences = "";

		returnSentences = removeRatios(returnIndexes, returnSentences);
			//comment out above line if you want to keep the passages that have ratios
			//also need to uncomment the line below
		//returnSentences = withRatios(returnIndexes, returnSentences);  //creating the three 
sentence passage

		return returnSentences;
	}

	//returning an arraylist with all the indices with sentences that have property search terms
	private ArrayList<HashSet<Integer>> 
getPropertyMatchingIndexes(ArrayList<ArrayList<String>> searchTermList, 
ArrayList<ArrayList<String>> exclusionTermList)
	{
		ArrayList<HashSet<Integer>> allMatchingIndexes = new 
ArrayList<HashSet<Integer>> ();
		int counter = 0;
		for(ArrayList<String> searchTerms : searchTermList)
		{
			allMatchingIndexes.add(getMatchingIndexesWithExclude(searchTerms, 
exclusionTermList.get(counter)));
			counter++;
		}
		return allMatchingIndexes;
	}

	//identifies indices of sentences containing all properties selected (for multiple properties)
	//for one property, it's just creating a HashSet with the indices
	//for multiple properties, it's creating a HashSet with the indices that overlap
	private HashSet<Integer> multiPropertyHash(ArrayList<HashSet<Integer>> matchingIndexes)
	{
		HashSet<Integer> combinedIndexes = new HashSet<Integer> (matchingIndexes.get(0));
		for(int i = 1; i < matchingIndexes.size(); i++)
		{
			HashSet<Integer> tempIndexes = new HashSet<Integer> 
(matchingIndexes.get(i));
			combinedIndexes.retainAll(tempIndexes);
		}
		return combinedIndexes;
```

```
	}

	//for stand-alone terms
	//identifies indices of sentences containing all properties selected (for multiple properties)
	//for one property, it's just creating a HashSet with the indices
	//for multiple properties, it's creating a HashSet with the indices that overlap
	private HashSet<Integer> multiSAPropertyHash(ArrayList<HashSet<Integer>>
matchingSAIndexes, ArrayList<HashSet<Integer>> matchingRegIndexes)
	{
		HashSet<Integer> combinedIndexes = new HashSet<Integer> ();
		HashSet<Integer> temp = new HashSet<Integer> ();
		HashSet<Integer> tempCombined = new HashSet<Integer> ();

		if (matchingSAIndexes.size()==1)
			combinedIndexes = matchingSAIndexes.get(0);
		else
		{
			for(int j = 0; j < matchingSAIndexes.size(); j++)
			{
				temp = matchingSAIndexes.get(j);
			for(int i = 0; i < matchingRegIndexes.size(); i++)
				{
				if (i!=j)
				{
						HashSet<Integer> tempIndexes = new
HashSet<Integer> (matchingSAIndexes.get(i));
						tempIndexes.addAll(matchingRegIndexes.get(i));
						temp.retainAll(tempIndexes);
						tempCombined = temp;
				}
				combinedIndexes.addAll(tempCombined);
				}
			}
		}
		return combinedIndexes;
	}

	//returning the indices of the sentences that have search terms, plus those before and after that fall
within the range
	//using for modification, causation, comparison, correlation terms
	//		if notPattern = 0, then search as normal
	//		if notPattern = 1, then check for not+phrase by using getMatchingIndexes and
matchNotPhrases
	private HashSet<Integer> getHelperIndexes(String[] phrases, int range, int notPattern)
	{
		HashSet<Integer> returnIndexes = new HashSet<Integer> ();
		for(Integer matchIndex : getMatchingIndexes(new
ArrayList<String>(Arrays.asList(phrases)), notPattern))
		{
			for(int j = 0; j <= range; j++)
			{
				if (j == 0)
						returnIndexes.add(matchIndex);
```

```
                    else
                        {
                        returnIndexes.add(matchIndex+j);
                        if (matchIndex-j >=0)
                            returnIndexes.add(matchIndex-j);
                        }
                }
            }
            return returnIndexes;
        }

        //find indexes of sentences containing "if" and "then" in the same sentence
        //then add relationshipRange number of sentences before and after
        //return indexes
        private HashSet<Integer> getIfThenIndexes(int range, int notPattern)
        {
            String[] ifPhrase= {"if"};
            String[] thenPhrase = {"then"};
            HashSet<Integer> ifIndexes = getMatchingIndexes(new
    ArrayList<String>(Arrays.asList(ifPhrase)), notPattern);
            HashSet<Integer> thenIndexes = getMatchingIndexes(new
    ArrayList<String>(Arrays.asList(thenPhrase)), notPattern);
            HashSet<Integer> relationshipIfThenIndexes = getMergedIndexes(ifIndexes,
    thenIndexes);

            HashSet<Integer> returnIndexes = new HashSet<Integer> ();

            for (int i = 0; i < sentences.size(); i++)
            {
                if (relationshipIfThenIndexes.contains(i))
                {
                    for (int j = 0; j<=range; j++)
                    {
                        if (j==0)
                            returnIndexes.add(i);
                        else
                        {
                            returnIndexes.add(i+j);
                            if (i-j >= 0)
                                returnIndexes.add(i-j);
                        }
                    }
                }
            }
            return returnIndexes;
        }

        //merge the modification indexes and the property indexes (intersection of two sets)
        private HashSet<Integer> getMergedIndexes(HashSet<Integer> propertyIndexes,
    HashSet<Integer> identifierIndexes)
        {
            HashSet<Integer> returnIndexes = new HashSet<Integer>(propertyIndexes);
            returnIndexes.retainAll(identifierIndexes);
```

```
		return returnIndexes;
	}

	//returning the indices of the sentences that have search terms
	private HashSet<Integer> getMatchingIndexes(ArrayList<String> searchTerms, int notPattern)
	{
		HashSet<Integer> matchingIndexes = new HashSet<Integer> ();
		if (notPattern==0)
		{
			for(int index = 0; index < sentences.size(); index++)
			{
				if(sentences.get(index).matchPhrases(searchTerms))
				{
						matchingIndexes.add(index);
				}
			}
		}

		if (notPattern==1)
		{
			for(int index = 0; index < sentences.size(); index++)
			{
				if(sentences.get(index).matchPhrasesNot(searchTerms))
				{
						matchingIndexes.add(index);
				}
			}
		}
		return matchingIndexes;
	}

	//returning the indices of the sentences that have search terms
	//adding in the use of exclusion terms
	private HashSet<Integer> getMatchingIndexesWithExclude(ArrayList<String> searchTerms,
ArrayList<String> exclusionTerms)
	{
		HashSet<Integer> matchingIndexes = new HashSet<Integer> ();
		for(int index = 0; index < sentences.size(); index++)
		{
			if(sentences.get(index).matchPhrasesWithExclude(searchTerms,
exclusionTerms))
			{
				matchingIndexes.add(index);
			}
		}
		return matchingIndexes;
	}

	//checking for number pattern in the property sentence
	//if found, remove sentence index from returnIndexes
	private TreeSet<Integer> removeNumbers(TreeSet<Integer> returnIndexesInitial)
	{
		ArrayList<Integer> returnIndexesList = new ArrayList<Integer>(returnIndexesInitial);
```

```
for(int i = 0; i < returnIndexesList.size(); i++)
{
    String sentence = "";
    sentence += sentences.get(returnIndexesList.get(i));

    Pattern p = Pattern.compile("\\b-?\\d+\\b");
    Matcher m = p.matcher(sentence);
    HashSet<Integer> check = new HashSet<Integer> ();
    while (m.find())
    {
        if (m.end()+7 < sentence.length() && sentence.substring(m.end()+1,
m.end()+8).contains("percent"))
        {
            check.add(0);
        }
        else if (m.end()+8 < sentence.length() &&
(sentence.substring(m.end(), m.end()+9).contains("per cent") || sentence.substring(m.end(),
m.end()+9).contains("percent")))
        {
            check.add(0);
        }
        else if (m.end()-m.start()==4 && (sentence.substring(m.start(),
m.start()+2).contains("19") || sentence.substring(m.start()+0, m.start()+2).contains("20")))
        {
            check.add(0);
        }
        else if (m.start()-3 >= 0 && sentence.substring(m.start()-3,
m.start()).contains("eq"))
        {
            check.add(0);
        }
        else if (m.start()-4 >= 0 && sentence.substring(m.start()-4,
m.start()).contains("fig"))
        {
            check.add(0);
        }
        else if (m.start()-6 >= 0 && sentence.substring(m.start()-6,
m.start()).contains("table"))
        {
            check.add(0);
        }
        else if (m.start()-7 >= 0 && sentence.substring(m.start()-7,
m.start()).contains("figure"))
        {
            check.add(0);
        }
        else if (m.start()-9 >= 0 && sentence.substring(m.start()-9,
m.start()).contains("equation"))
        {
            check.add(0);
        }
        else if (sentence.substring(0, m.start()).length() < 23 &&
(sentence.substring(0, m.start()).contains("figs") || sentence.substring(0, m.start()).contains("figures")))
```

```
{
    check.add(0);
}
else if (m.start()-23 >= 0 && (sentence.substring(m.start()-23, m.start()).contains("figs") || sentence.substring(m.start()-23, m.start()).contains("figures")))
{
    check.add(0);
}
else if (sentence.substring(0, m.start()).length() < 14 && (sentence.substring(0, m.start()).contains("eqs")))
{
    check.add(0);
}
else if (m.start()-14 >= 0 && (sentence.substring(m.start()-14, m.start()).contains("eqs")))
{
    check.add(0);
}
else if (sentence.substring(0, m.start()).length() < 25 && (sentence.substring(0, m.start()).contains("equations")))
{
    check.add(0);
}
else if (m.start()-25 >= 0 && (sentence.substring(m.start()-25, m.start()).contains("equations")))
{
    check.add(0);
}
else if (sentence.substring(0, m.start()).length() < 14 && (sentence.substring(0, m.start()).contains("tables")))
{
    check.add(0);
}
else if (m.start()-14 >= 0 && (sentence.substring(m.start()-14, m.start()).contains("tables")))
{
    check.add(0);
}
else if (sentence.substring(m.end()).length() < 15 && sentence.substring(m.end()).contains("percent")) //checking for x percent
{
    check.add(0);
}
else if (m.end()+16 < sentence.length() && sentence.substring(m.end(), m.end()+16).contains("percent")) //checking for x percent
{
    check.add(0);
}
else
{
    check.add(1);
}
}
```

```
				if (check.contains(1))
				{
					returnIndexesList.remove(i);
					i=i-1;
				}
			}

			TreeSet<Integer> returnIndexesFinal = new TreeSet<Integer>(returnIndexesList);
			return returnIndexesFinal;
		}

		//checking for ratios
		//also creating the three sentence passages for the article
		private String removeRatios(TreeSet<Integer> returnIndexes, String returnSentences)
		{
			for(Integer i : returnIndexes)
			{
				String passage = "";
				if (i-1>=0)
					passage += sentences.get(i-1) + "\n";
				passage += sentences.get(i)+ "\n";
				if (i+1<sentences.size())
					passage += sentences.get(i+1)+ "\n";

				String fullPassage = passage;
				HashSet<Integer> check = new HashSet<Integer> ();

				if (passage.contains("/"))
				{
					int index = passage.indexOf("/");
					while (index != -1)
					{
						if (index-1>=0 && index+1<passage.length())
						{
							if (!(Character.isDigit(passage.charAt(index-
1))==true) || !(Character.isDigit(passage.charAt(index+1))==true))
							{
								if (index-3>=0 &&
index+4<=passage.length())
								{
									String before =
passage.substring(index-3, index);
									String after =
passage.substring(index+1, index+4);
									if ((before.matches("[a-
z]+")==true) || (after.matches("[a-z]+")==true))
										check.add(0);
									else
									{
										check.add(1);
									}
								}
```

```java
									else
									{
										check.add(1);
									}
								}
								else
								{
									check.add(1);
								}
							}
							else
							{
								check.add(1);
							}
							passage = passage.substring(index + 1);
							index = passage.indexOf("/");
						}
					}

					if (!check.contains(1) || check.isEmpty())
					{
						returnSentences += fullPassage;
						returnSentences += "\n";
					}
				}
				return returnSentences;
			}

			//Creating the three sentence passage without removing ratios (/)
			private String withRatios(TreeSet<Integer> returnIndexes, String returnSentences)
			{
				for(Integer i : returnIndexes)
				{
					if (i-1>=0)
						returnSentences += sentences.get(i-1) + "\n";
					returnSentences += sentences.get(i) + "\n";
					if (i+1<sentences.size())
						returnSentences += sentences.get(i+1) + "\n";
					returnSentences += "\n";
				}
				return returnSentences;
			}
}
```

### Sentence Class

```java
import java.util.ArrayList;
import java.util.HashSet;

public class Sentence {
	private ArrayList<String> wordList;
	private HashSet<String> wordHash;
```

```
	public Sentence()
	{
		wordList = new ArrayList<String>();
		wordHash = new HashSet<String>();
	}

	public void addWord(String word)
	{
		wordList.add(word.toLowerCase());
		wordHash.add(word.toLowerCase());
	}

	//returns true if the sentence itself contains any word from the ArrayList of phrases
	public boolean matchPhrases(ArrayList<String> phrases)
	{
		String s = getSentence();
		for(String phrase : phrases)
		{
			if (wordHash.contains(phrase) || (phrase.contains(" ") && s.contains(phrase)))
			{
				return true;
			}
		}
		return false;
	}

	//returns true if the sentence itself contains any word from the ArrayList of phrases (minus those
words in the exclusion terms)
	//adding in the exclusion terms
	public boolean matchPhrasesWithExclude(ArrayList<String> phrasesSearch, ArrayList<String>
phrasesExclude)
	{
		String s = getSentence();
		ArrayList<Boolean> goodSentenceExclusion = new ArrayList<Boolean> ();
		for(String phrase : phrasesSearch)
		{
			if (wordHash.contains(phrase) || (phrase.contains(" ") && s.contains(phrase)))
			{
				Boolean goodSentence = true;
				for(String exclude : phrasesExclude)
				{
					if (exclude.contains(phrase) && s.contains(exclude))
						goodSentence = false;
				}
				goodSentenceExclusion.add(goodSentence);
			}
		}
		if (goodSentenceExclusion.contains(true))
			return true;
		else
			return false;
	}
```

```
	//returns true if the sentence itself contains any word from the ArrayList of phrases
	//except if there is "not"+phrase
	//this is an irrelevant passage pattern
	public boolean matchPhrasesNot(ArrayList<String> phrases)
	{
		String s = getSentence();
		HashSet<Integer> check = new HashSet<Integer> ();
		for(String phrase : phrases)
		{
			if (wordHash.contains(phrase) || (phrase.contains(" ") && s.contains(phrase)))
			{
				if (s.contains("not "+phrase))
					check.add(1);
				else
					check.add(0);
			}
		}
		if (check.contains(0))
			return true;
		else
			return false;
	}

	public String getSentence()
	{
		return wordList.toString().replaceAll("[\\[\\],]", "");
	}

	public String toString()
	{
		return getSentence();
	}

	public int length()
	{
		return wordList.size();
	}
}
```

# APPENDIX J

## STAGE 4 ALGORITHM OUTPUT FOR HARDNESS

Article Info:
i1 Extreme strength observed in limpet teeth
Asa H. Barber, Dun Lu, and Nicola M. Pugno
Accepted: 23 January 2015
Search Result:
the goethite nanofibres are expected to dictate the flaw tolerance of the resultant composite owing to their diameters being below a critical threshold value of the order of tens of nanometres
this work demonstrates a high strength composite found in nature and highlights a design strategy towards strong engineered composites reinforced with a high volume fraction of nanofibrous material as the limpet tooth is effective at resisting failure owing to abrasion as demonstrating during rasping of the tooth over rock surfaces corresponding structural design features are expected to be significant for novel biomaterials with extreme strength and hardness such as next generation dental restorations

Article Info:
i3 Isotropic microscale mechanical properties of coral skeletons
Luca Pasquini1, Alan Molinari1, Paola Fantazzini1, Yannicke Dauphen2, Jean-Pierre Cuif2, Oren Levy3, Zvy Dubinsky3, Erik Caroselli4, Fiorella Prada4, Stefano Goffredo4, Matteo Di Giosia5, Michela Reggi5 and Giuseppe Falini5
Accepted: 18 March 2015
Search Result:
the influence of porosity and organic matrix on the mechanical properties can also explain the lower youngs modulus of coral skeletons with respect to polycrystalline aragonite
similarly the difference in hardness between the two coral species may reflect a lower content of organic matrix and possibly porosity in s
pistillata with respect to b

pistillata are in the ranges 76 77 gpa and 4951 gpa respectively
the statistical analysis does not reveal a significant difference between skeletal sections having different orientations with respect to the main growth axis of the coral the youngs modulus is the same for the two coral species whereas hardness is approximately 3 percent lower in b
the mechanical properties are also rather constant over different indentation sites typically separated by few tens of micrometres within one section

orientation dependence of the mechanical properties of the entire skeleton if present should therefore be ascribed to its anisotropic shape on a larger ie mm to cm scale
the slightly lower youngs modulus by about 10 percent with respect to pure polycrystalline aragonite can be ascribed to the presence of microporosity and soft organic matrix in the coral skeletons
in comparison the widely studied nacre material exhibits a clear anisotropy of youngs modulus and a significant variation among different seashell species

Article Info:
i9 Compliant threads maximize spider silk connection strength and toughness
Avery Meyer1, Nicola M. Pugno2,3 and Steven W. Cranford1

Accepted: 16 June 2014
Search Result:
in webs as in engineered structures brittle like small deformation failure should be avoided
finally with the addition of pre stretch we note a distinct transition from a nonlinear stiffening response to a more softening like response
the hyperelastic character of the global response changes from an extreme stiffening like behaviour at low pre strains most similar to the constitutive law figure 2 to a near constant stiffness with sudden yield at high pre strains

stiffening behaviour in the anchorage is desired owing to the relative initial compliance
as the material transitions from linear elastic to softening behaviour there is little change in the connection capacity for the strain regime explored
this can be attributed relative stiffness of a plastic softening like behaviour at small strains for the case here where 1ult 2 at 1 015 the ratio of anchor thread tangent stiffness with b 05 to adhered thread tangent stiffness with b 3 is greater than 100 see methods section

Article Info:
i10 Lee N, Horstemeyer MF, Rhee H, Nabors B, Liao J, Williams LN. 2014 Hierarchical multiscale structure–property relationships of the red-bellied woodpecker (Melanerpes carolinus) beak. J. R. Soc. Interface
11: 20140274. http://dx.doi.org/10.1098/rsif.2014.0274
Search Result:
to examine the gradient of the nanomechanical properties experiments were performed on four different beak locations from the tip to the root and the results are depicted in figure 8
the results show a decrease in the hardness as the location changes from the tip to the root of the beak for the rhamphotheca
the bony layer did not show a change with respect to location

micromechanically the microhardness of the rhamphotheca of the woodpecker is about 50 percent greater than those of other birds such as the toucan hornbill and starling
it is reported that dark coloured beaks have a greater hardness than light coloured beaks so the dark beaks are less susceptible to wear as the colour of woodpeckers beaks is predominantly black then one would expect a greater hardness if the european starlings study is consistent with woodpeckers beaks
our study also shows that the microhardness of the core part of the woodpeckers beaks is indeed two to three times greater than those of the toucan and hornbill beaks table 1

Article Info:
i13 De Tommasi D, Millardi N, Puglisi G, Saccomandi G. 2013 An energetic model for macromolecules unfolding in stretching experiments. J R Soc Interface 10: 20130651.
http://dx.doi.org/10.1098/rsif.2013.0651
Search Result:
these transitions have a main role in the energetics of the chain unfoldings owing to the two following effects first there is the enthalpic contribution of the transition itself debonding energy then there is an entropic contribution associated with the variation of the microstructure leading to increased contour lengths of the chains
in the final stage leading to the material failure the behaviour is regulated by the entropic hardening of the fully unfolded macromolecules primary structure
to fix the ideas in figure 1 we schematically show a typical afm single molecule stretching experiment reproduced from on an engineering reconstructed macromolecule

in the following section we discuss this issue and propose an extension of our model able to reproduce the hardening effect observed eg in titin macromolecule unfolding
unfolding energy hierarchy the described variable observations on subsequent transition forces force plateaux hardening softening have been given different physical interpretations
a hierarchy of unfolding energies of the unfolding crystals may be simply due to inhomogeneity effects of the crystal domains showing variable bond breaking barriers possibly owing to interfacial energy effects another important effect is anisotropicity of the crystals with respect to the force directiondifferent paths in the wiggly energy landscape lead to different unfolding energy barriers so that different unfolding forces can be induced by variable crystal orientations in the macromolecule

it is important to remark that in the case of increasing unfolding energies the di block approximation can fail with the order of unfolding that is regulated by the competition of mixing energy interfacial energy effects and variability of the unfolding energy
this competition regulates in the quasi static regime the order of transition and the hardening softening or non monotonic law of the successive unfolding force
aimed again at a fully analytical result we here suppose that both the mixing entropy contribution and the interfacial energy term are negligible when compared with the unfolding energy increment dq so that the unfolding evolution strategy is regulated by the unfolding energy hierarchy and we may again describe the stretching behaviour of the macromolecule by minimizing an energy as in 36 but with variable unfolding energy increments

Article Info:
i17 Epidermis architecture and material properties of the skin of four snake species
Marie-Christin G. Klein* and Stanislav N. Gorb
Published online 15 August 2012
Search Result:
the presence of such a gradient is a possible adaptation to locomotion and wear minimization on natural substrates
in general the difference in both the effective elastic modulus and hardness of the osl and isl between species was not large compared with the difference in epidermis thickness and architecture
introduction snakes are limbless reptiles that use their entire body for sliding locomotion

it is suggested that both the spacing and dimensions of these structures play an important role in the ability to detect and enhance the absorption of infrared radiation in sliding contacts local stress concentrations on the surface lead to material fatigue and failure
the surface is more effective against abrasion wear if it has a gradient in material structure and properties because it leads to a more uniform stress distribution and thus to the minimization of local stress concentrations a hard inflexible surface material easily forms cracks under pressure whereas a soft flexible system will be easily worn off under shear stress
as we have previously assumed the gradient in material properties from a stiff surface to a soft depth will improve wear resistance in snake skin by combining the advantages of stiffness and flexibility

the modulus values for the isls exhibit a significant difference between the outer and inner most intervals 300400 and 15001600 nm and not between the others
this implies a smooth transition from the soft to the harder regions
the hd values on the other hand for the inner scale values are significantly different except for the intervals 10001100 and 15001600 nm which implies that the interval 300400 nm is very soft in comparison with the others and that the transition is not very smooth

discussion our morphological results and nanoindenation data obtained on the ventral scales provide strong evidence for the existence of a gradient in material properties in the four snake species studied
the osls are harder and exhibit a higher eem than the inner ones
the results of this study give good support to our previous data on g

Article Info:
i19 Cell wall structure and formation of maturing fibres of moso bamboo (Phyllostachys pubescens) increase buckling resistance
Xiaoqing Wang1,*, Haiqing Ren1, Bo Zhang2,3, Benhua Fei2 and Ingo Burgert3,*
Published online 14 September 2011
Search Result:
the lower stiffness of the inner cell wall region is likely to be an artefact because of indenting close to the lumen of the fibre
it has been shown that the indentation modulus and hardness are highly dependent on both cellulose microfibril angle and lignin level of plant cell walls however the degree of lignification seems to have a minor influence on the cell wall stiffness of bamboo fibres which can be explained by the almost longitudinal orientation of the cellulose fibrils owing to lack of a vascular cambium arborescent monocotyledons show no secondary thickening growth that impedes geometrical adaptations to mechanical loads and increases the necessity of structural optimization at the material level
one crucial aspect for the mechanical stability of the trunk is the proper embedding of the stiff fibre caps of the vascular bundles into the soft parenchymatic matrix

in a previous study on a specific fibre cap type of the palm w
robusta a gradual decrease in stiffness across the fibre capwas observed to avoid stress discontinuities between the stiffening elements and the soft parenchyma tissues which was interpreted as an adaptation to the required mechanical constraints under the given growth conditions the stiffness gradient in the fibre cap of the palm is developed at two hierarchical levels on the one hand at the tissue level owing to a decrease of tissue density cell wall thickness/cell size from the inner to the outer region and on the other hand at the level of the cell wall assembly by regulating the level of lignin and its composition
in the case of the palm fibre cap this affects the axial tensile stiffness because the cell walls possess a high microfibril angle and therefore the axial material properties of cell walls are highly dependent on the shear stiffness and strength of the matrix in terms of the investigated bamboo culm the results do not indicate changes in cell wall properties across the fibre cap

Article Info:
i20
Evolutionary optimization of material properties of a tropical seed
Peter W. Lucas1,*, John T. Gaskins2, Timothy K. Lowrey3, Mark E. Harrison4,5, Helen C. Morrogh-Bernard5, Susan M. Cheyne5,6 and Matthew R. Begley7
Published online 25 May 2011
Search Result:
the shell has a nearly spatially uniform hardnessthe highest recorded for nut shells
while the elastic modulus varies more than hardness the major influence is whether loading on the fibres is axial or transverse table 1
clearly in zone ii the mean modulus is a statistical average over a range of orientations

analysis of large animal predation orangutans pongo pygmaeus large fruit eating great apes whose feeding behaviour has been studied intensively eat m

parviflora seeds at a number of study sites in borneo and sumatra in sabangau central kalimantan adult orangutans spent a mean 133182 percent of their total monthly feeding time consuming these seeds which are the hardest item they are known to consume observations of orangutan feeding suggest two mechanisms for large animal predation ie two methods to crack the seed by biting
in the first method observed in wild orangutans in sabangau the seed is placed between jaws with the germination band running vertically from bottom tooth to top tooth the crack probably originates from the plug which serves as a pre crack aligned with the bite force and pointing into the germination band figure 4

the calculation of internal pressure required to crack the shell suggests that shell thickness and stiffness are maximal thicker stiffer shells would require pressures that significantly exceed known turgor pressures preventing germination
while softer more compliant shells would facilitate germination by requiring lower turgor pressures to crack the shell open even small reductions in hardness and stiffness would leave the shell prone to burrowing by beetles
similarly the fracture toughness of the shell is low enough to permit germination but not sufficiently low as to allow chipping by rodents

to resist the attack of large animals the shell needs to be thick stiff and most of all tough rather than maximizing these properties outright the shell has evolved a complex structure that raises these properties to upper limit indicated by the need to crack open during germination
the calculations outlined above emphasize the critical role of the germination band it is equally as hard as the surrounding shell which is the principal property hindering burrowing by very small predators
at the same time this band has a lower fracture toughness to allow for germination while being narrow enough to prevent chipping by medium sized predators

# APPENDIX K

## REVISED STAGE 1 ALGORITHM OUTPUT FOR HARDNESS

Article Info:
i1 Extreme strength observed in limpet teeth
Asa H. Barber, Dun Lu, and Nicola M. Pugno
Accepted: 23 January 2015
Search Result:
introduction composite structures are widespread in nature and are ubiquitous in mineralized tissue where protein based polymer frameworks are reinforced with a stronger and stiffer mineral phase these composite structures often have a distinct mechanical function and have led to a number of engineering principles being applied to explain resultant structurefunction behaviour in biological organisms more recent concepts have examined the potential of biology in controlling the size of constituents in natural composite structures particularly for enhanced mechanical properties at small length scales
specifically the reinforcing mineral phase in many organisms approaches nanometre length scales at least in one dimension which has been proposed as promoting flaw insensitivity to increase the tensile strength of mineralized tissue the enhancement of material tensile strength owing to size effects has additionally been shown historically including griffiths observations of increased glass fibre failure stress as their diameters decreased to more recent quantized fracture mechanics qfm extensions from statistical descriptions of material strength by weibull however the insensitivity of materials to flaws has been shown to operate at length scales of many tens of nanometres so that material failure is governed by the theoretical strength of the material and not by stress concentrations around flaws as first considered by griffith discrete examples of exceptional strength in natural materials are perhaps most prevalent in the silk of spiders with tensile strength values of up to 45 gpa recorded in the literature limpet teeth shown in figure 1 are an example of a material produced biologically that is optimized for strength especially as these teeth need to be extremely strong and hard to avoid catastrophic failure when rasping over rock surfaces during feeding
recent work has shown that the teeth of limpets approximate to an almost ideal model natural composite material where high aspect ratio mineral nanofibres of goethite reinforce a protein matrix limpet teeth are also notable by displaying a lack of structural hierarchy present in many other mineralized tissue structures

the goethite nanofibres are expected to dictate the flaw tolerance of the resultant composite owing to their diameters being below a critical threshold value of the order of tens of nanometres
this work demonstrates a high strength composite found in nature and highlights a design strategy towards strong engineered composites reinforced with a high volume fraction of nanofibrous material as the limpet tooth is effective at resisting failure owing to abrasion as demonstrating during rasping of the tooth over rock surfaces corresponding structural design features are expected to be significant for novel biomaterials with extreme strength and hardness such as next generation dental restorations

Article Info:
i3 Isotropic microscale mechanical properties of coral skeletons
Luca Pasquini1, Alan Molinari1, Paola Fantazzini1, Yannicke Dauphen2, Jean-Pierre Cuif2, Oren Levy3, Zvy Dubinsky3, Erik Caroselli4, Fiorella Prada4, Stefano Goffredo4, Matteo Di Giosia5, Michela Reggi5 and Giuseppe Falini5
Accepted: 18 March 2015
Search Result:
in this work the skeletons of two coral species solitary balanophyllia europaea and colonial stylophora pistillata were investigated by nanoindentation

the hardness hit and youngs modulus eit were determined from the analysis of several loaddepth data on two perpendicular sections of the skeletons longitudinal parallel to the main growth axis and transverse within the experimental and statistical uncertainty the average values of the mechanical parameters are independent on the sections orientation

the homogeneous and isotropic mechanical behaviour of the coral skeletons at the microscale is correlated with the microstructure observed by electron microscopy and atomic force microscopy and with the x ray diffraction patterns of the longitudinal and transverse sections
introduction scleractinian corals represent a major source of biogenic calcium carbonate and are among the fastest marine mineralizing organisms their skeleton is a composite structure with both inorganic aragonite and organic components the content of organic components and structural water ranges between 1 and 3 wt% whereas non structural water represents a minor component being present in amounts lower than 05 wt% one of the most important roles of coral skeletons is the building of the structure on which the soft tissue can grow and be protected
the skeletal structure of the corals also make the framework of the reef which has an important ecological economical and social relevance a detailed description of corals skeletal texture and morphogenesis is reported in several reviews eg and references therein

nanoindentation techniques have recently been adapted for the study of biological materials and are a powerful tool for study of the mechanical properties at the nano microscale
the analysis of the platelets on the nacreous layer of the red abalone shell showed that the deformability of the aragonite platelets together with the crack deflection aragonite platelet slip and organic adhesive interlayer contribute to the nacres fracture toughness sea urchin spines from heterocentrotus mammillatus phyllacanthus imperialis and prionocidaris baculosa showed a strong dependence of the indentation modulus but not the indentation hardness on the local porosity
this was attributed to the network type of porosity the hardness and modulus of biogenic calcite from the prismatic layer of the mollusc atrina rigida was compared with a pure geological calcite iceland spar

the analysis of the platelets on the nacreous layer of the red abalone shell showed that the deformability of the aragonite platelets together with the crack deflection aragonite platelet slip and organic adhesive interlayer contribute to the nacres fracture toughness sea urchin spines from heterocentrotus mammillatus phyllacanthus imperialis and prionocidaris baculosa showed a strong dependence of the indentation modulus but not the indentation hardness on the local porosity
this was attributed to the network type of porosity the hardness and modulus of biogenic calcite from the prismatic layer of the mollusc atrina rigida was compared with a pure geological calcite iceland spar
on the 001 face biogenic calcite was found to be 5070% harder than geologic calcite

on the 001 face biogenic calcite was found to be 5070% harder than geologic calcite
the higher hardness and increased anisotropy of biogenic calcite was accounted for by hardening mechanisms based on hindered dislocation motion rather than on crack deflection the mechanical properties of the nacreous layer of five different seashells were investigated by nanoindentation and three point bending tests it was found that the aspect ratio of the mineral phase in all seashells is close to the optimal value for strength as predicted by theory the multiscale mechanical properties of nacre from the single aragonite platelet to the composite brick and mortar structure were studied with great care using a combination of spherical and sharp nanoindentation tests
the elastic properties of the intracrystalline organic phase and its role in the deformation of the aragonite platelet were elucidated the compressive strength of acropora sp

care was also taken to keep a minimum distance of 50 mm from the coral edges and from microscale pores visible on the surface
this is important in order to avoid artefacts owing to the presence of elastic discontinuities in proximity of the indentation area

if the indentation areas on the two perpendicular sections are properly selected this procedure amounts to indenting a small volume less than 1 mm3 of the original skeleton along two perpendicular directions

if the indentation areas on the two perpendicular sections are properly selected this procedure amounts to indenting a small volume less than 1 mm3 of the original skeleton along two perpendicular directions
figure 1 also displays videomicroscope images of typical indentation zones for b
europaea d and s

the measurements were conducted in load control mode using the following settings maximum load 50 mn loading/ unloading rate 100 mn min 1 holding period at maximum load 10 s
in addition for each section one indentation was performed in dynamic sinus mode in which an oscillation at 5 hz and 10% amplitude was superimposed to the rising load
the instrumented youngs modulus eit and hardness hit were determined by the oliver pharr op method hit is given by the ratio between the maximum applied load and the corresponding projected contact area whereas eit is derived from the initial slope of the loaddisplacement curve during unloading

in addition for each section one indentation was performed in dynamic sinus mode in which an oscillation at 5 hz and 10% amplitude was superimposed to the rising load
the instrumented youngs modulus eit and hardness hit were determined by the oliver pharr op method hit is given by the ratio between the maximum applied load and the corresponding projected contact area whereas eit is derived from the initial slope of the loaddisplacement curve during unloading
the dynamic analysis of sinus mode measurements permits determination of hit and eit as a function of the penetration depth

the instrumented youngs modulus eit and hardness hit were determined by the oliver pharr op method hit is given by the ratio between the maximum applied load and the corresponding projected contact area whereas eit is derived from the initial slope of the loaddisplacement curve during unloading
the dynamic analysis of sinus mode measurements permits determination of hit and eit as a function of the penetration depth
x ray diffraction scanning electron microscopy thermogravimetrical analyses and atomic force microscopy measurements xrd profiles of each section were collected using an x celerator powder diffractometer panalytical using cu ka radiation 1540 a

pistillata was of 22 01 wt% and 1401 wt% respectively from tga
youngs modulus and hardness of coral skeletons figures 4 and 5 show the loaddepth nanoindentation curves measured on a single dry skeleton of the two coral species
in figures 4 and 5ab it is possible to note that a relatively small variability exists between tests carried out on the same section

species and orientation were taken as the between subjects factor and within subjects factor respectively this statistical analysis of skeletal mechanical properties clearly shows that i there is no significant difference between the longitudinal and transverse sections this is true for both species ii in the case of youngs modulus eit there is no significant difference between the two species iii in the case of hardness hit there is a small yet statistically significant difference between the two species with s
pistillata being about 3% harder than b

pistillata being about 3% harder than b
the dynamic sinus measurements reported in electronic supplementary material in figures s2s3 for the longitudinal sections of the two species demonstrate that the corals mechanical properties do not depend significantly on the indentation depth in the probed range
the same conclusion applies to the transverse sections of the two corals

the anisotropy becomes even more evident in single crystal aragonite figure 6b where the average nanoindentation curves for the 001 and 122 sections exhibit a remarkable difference both in the residual depth and in the unloading slope

according to the op analysis reported in table 1 nacre exhibits a strong eit anisotropy and a weak hit anisotropy whereas for single crystal aragonite both eit and hit strongly depend on the indentation direction

the corals mechanical parameters eit 7677 gpa and hit 4951 gpa are higher than those of the nacreous layer in a

rigida and lower than those of the stiff and hard 001 direction of single crystal aragonite

they appear close to the values determined for the compliant and soft 122 direction

residual indents and crack generation typical optical images of the residual indents on the corals skeletons recorded in situ right after the measurements are displayed in electronic supplementary material figure s4

the side of the residual indentation triangle is 5 mm

in some cases the surface around the indent edges appears very flat electronic supplementary material figure s4ac whereas in others a disturbance of the surface is observed electronic supplementary material figure s4bd

higher magnification sem images of these residual indents clearly show the occurrence of the pile up of material as a result of the plastic deformation

the occurrence of pile up and microcracks was examined in more detail for b

europaea by performing further groups of 10 indentations with maximum loads of 5 and 500 mn

at 5 mn maximum load we hardly observed any pile up formation or microcrack generation

the length of the crack enables the determination of another interesting mechanical property ie the fracture toughness kc according to a described procedure 22 and references therein

even though a precise estimation of kc is difficult owing to the variability of crack lengthwe can assert that also the fracture toughness of the corals skeletons is independent on the indentation direction kc 0601 mpa m1/2 and kc 0501 mpa m1/2 for the longitudinal and transverse sections respectively

conversely no microcrack generation was observed in nacre even at the highest load of 500 mn indicating a higher fracture toughness of nacre with respect to the corals skeletons

the microscopic origin of this behaviour is the existence of preferential slip systems such as the f110g001 family the pile up effect around the indentations was described correctly by a crystal plasticity model which takes into account all the slip systems our values table 1 are in good agreement with reference and confirm the strong anisotropy of the nanoindentation hardness

the slightly higher hardness measured here for the section 001 could be attributed to a higher concentration of impurities in our geogenic aragonite acting as pinning centres for dislocations

family the pile up effect around the indentations was described correctly by a crystal plasticity model which takes into account all the slip systems our values table 1 are in good agreement with reference and confirm the strong anisotropy of the nanoindentation hardness

the slightly higher hardness measured here for the section 001 could be attributed to a higher concentration of impurities in our geogenic aragonite acting as pinning centres for dislocations

anisotropic mechanical properties of nacre the mechanical properties of nacre and their relation to the materials microstructure are an intensively studied subject as mentioned in the introduction our motivation for measuring nacre was to achieve a direct comparison ie using the same equipment and protocol with the mechanical properties of coral skeletons

anisotropic mechanical properties of nacre the mechanical properties of nacre and their relation to the materials microstructure are an intensively studied subject as mentioned in the introduction our motivation

for measuring nacre was to achieve a direct comparison ie using the same equipment and protocol with the mechanical properties of coral skeletons
fleischli et al have shown that both hardness and youngs modulus strongly depend on the scale of nacres well known brick and mortar architecture especially on the thickness of the aragonite platelets
in the seashells with thick platelets in particular trochus maculatus and haliotis rufescens the values of eit and hit are close to pure aragonite conversely in the seashells characterized by thin platelets such as those of pteria penguin a youngs modulus as low as 608 gpa and a hardness of 3710 gpa were measured our values for the nacreous layer of a

fleischli et al have shown that both hardness and youngs modulus strongly depend on the scale of nacres well known brick and mortar architecture especially on the thickness of the aragonite platelets
in the seashells with thick platelets in particular trochus maculatus and haliotis rufescens the values of eit and hit are close to pure aragonite conversely in the seashells characterized by thin platelets such as those of pteria penguin a youngs modulus as low as 608 gpa and a hardness of 3710 gpa were measured our values for the nacreous layer of a
rigida table 1 fall within this range such reduced values are generally attributed to the high weight fraction around 6 wt% of organic matrix the mortar which is orders of magnitude softer and more compliant than aragonite the significant youngs modulus anisotropy in nacre highlighted by our measurements originates with the previously discussed anisotropy of the aragonite platelets the bricks which are aligned preferentially along one crystallographic direction within the nacreous layer

isotropic mechanical properties of coral skeletons the statistical analysis of nanoindentation data table 1 and electronic supplementary material tables s1 and s2 proves that coral skeletons in contradistinction to nacre exhibit isotropic mechanical behaviour at the microscale both in the elastic and plastic regimes despite the remarkable mechanical anisotropy of aragonite
in addition relatively small fluctuations in the range of about 3% for eit and 6% for hit were detected within the same section figures 4 and 5 indicating a homogeneous mechanical response on a spatial scale larger than the typical indentation volume
these features can be explained by the spatial arrangement of aragonite crystals as highlighted by sem and afm observations figures 2 and 3 and xrd profiles figure 3

these features can be explained by the spatial arrangement of aragonite crystals as highlighted by sem and afm observations figures 2 and 3 and xrd profiles figure 3
in fact sem shows that the aragonite fibrous crystals are about 200400 nm thick ie much thinner than the typical indentation size 510 mm and are oriented along different directions in a fan like arrangement each indentation therefore represents an average over many differently oriented crystals resulting in a homogeneous response across a section
the absence of preferential orientation of the aragonite crystals suggested by xrd figure 3 explains why the mechanical parameters are the same in the two perpendicular sections

this consideration probably also extends to the fracture behaviour even though further experiments are needed to better assess this point
effects of porosity and organic matrix on skeletal mechanical properties the small fluctuations observed within the same section may arise from local variations from one indentation site to another of one or more of the following items 1 average orientation of aragonite crystals although xrd reveals no preferential orientation over the whole section it is not possible to rule out small changes on the micrometre scale
2 pore content as shown by presser et al local porosity influences the mechanical properties determined by nanoindentation in particular lowering youngs modulus with respect to a fully dense material using time domain nuclear magnetic relaxometry we have shown that coral skeletons contain pores with sizes ranging from about 10 mm down to few tens of nanometres such pores may either be too small to be detected under the videomicroscope or be hidden beneath the surface and their varying concentration could contribute to the observed fluctuations

2 pore content as shown by presser et al local porosity influences the mechanical properties determined by nanoindentation in particular lowering youngs modulus with respect to a fully dense material using time domain nuclear magnetic relaxometry we have shown that coral skeletons contain pores with sizes ranging from about 10 mm down to few tens of nanometres such pores may either be too small to be detected under the videomicroscope or be hidden beneath the surface and their varying concentration could contribute to the observed fluctuations
3 abundance of the organic matrix stempfle et al have shown that both the intercrystalline and intracrystalline organic matrix characterized by low youngs modulus 63 and 38 gpa respectively play an important role in the elastic and plastic deformation of nacres aragonite platelets as already recalled it is expected that regions richer in organic matrix display lower youngs modulus and hardness
the influence of porosity and organic matrix on the mechanical properties can also explain the lower youngs modulus of coral skeletons with respect to polycrystalline aragonite

europaea 2201 wt%
however the difference in hardness is very small only 3% while fluctuations up to 100% between different seashell species have been reported making it difficult to ascribe it to a specific mechanism
the comparison between microscale mechanical properties of several coral species having diverse organic matrix content and porosity may be the subject of future studies

pistillata are in the ranges 76 77 gpa and 4951 gpa respectively
the statistical analysis does not reveal a significant difference between skeletal sections having different orientations with respect to the main growth axis of the coral the youngs modulus is the same for the two coral species whereas hardness is approximately 3% lower in b
the mechanical properties are also rather constant over different indentation sites typically separated by few tens of micrometres within one section

the statistical analysis does not reveal a significant difference between skeletal sections having different orientations with respect to the main growth axis of the coral the youngs modulus is the same for the two coral species whereas hardness is approximately 3% lower in b
the mechanical properties are also rather constant over different indentation sites typically separated by few tens of micrometres within one section
sem afm and xrd show that this behaviour originates from the random orientation of aragonite fibres and their forming spheroid particles which are much thinner than the indentation size

the mechanical properties are also rather constant over different indentation sites typically separated by few tens of micrometres within one section
sem afm and xrd show that this behaviour originates from the random orientation of aragonite fibres and their forming spheroid particles which are much thinner than the indentation size
despite the significant anisotropy of the building units aragonite crystals the coral skeletal material is thus homogeneous and isotropic as concerns the mechanical properties at the microscale

orientation dependence of the mechanical properties of the entire skeleton if present should therefore be ascribed to its anisotropic shape on a larger ie mm to cm scale
the slightly lower youngs modulus by about 10% with respect to pure polycrystalline aragonite can be ascribed to the presence of microporosity and soft organic matrix in the coral skeletons
in comparison the widely studied nacre material exhibits a clear anisotropy of youngs modulus and a significant variation among different seashell species

Article Info:
i4 Micro- and nano-structural details of a spider's filter for substrate vibrations: relevance for low-frequency signal transmission

Maxim Erko1, Osnat Younes-Metzler1, Alexander Rack2, Paul Zaslansky3, Seth L. Young4, Garrett Milliron1, Marius Chyasnavichyus4, Friedrich G. Barth5, Peter Fratzl1, Vladimir Tsukruk4, Igor Zlotnikov1 and Yael Politi1
Accepted: 7 January 2015
Search Result:
the sam analysis of this surface was performed under deionized water using 400 hz and 820 hz respectively
nanoindentation nanoindentation measurements were performed using ubi1 nano indenter hysitron minneapolis mn usa
the same samples were used as for sam

each measurement included 64 indents at the respective pad region
the values for the reduced elastic modulus er and for hardness h were obtained from the loaddisplacement curves according to the oliver and pharr method during measurements the temperature was kept at 24c well above the glass transition temperature of the epicuticle 3
results and discussion 31

the elaborate morphology of the pad and in particular the sub structure of the material of which it is made are discussed in more detail in the following
the distal end of the pad the distal end of the pad is extremely soft showing only 100 mpa for the reduced elastic modulus as measured in nanoindentation experiments under hydrated conditions
in agreement with this we have also found that this pad region is highly hydrated under natural conditions and the least sclerotized part of the pad

more study is needed in order to answer this question
the softness 100 mpa of the pads distal contact region with the tarsus increases the contact area with the tarsus and allows the largest deformation upon tarsal deflection while only comparatively small displacement of the pad is observed at the contact region with the slits
in this way substrate displacements of the order of up to 200 micron in accordance with barth geethabali are translated to small non destructive slit compressions of only few micrometres 3507 mm at 91 tarsal deflection and overstimulation or slit damage is avoided

Article Info:
i5 Nanomechanical properties of bird feather rachises: exploring naturally occurring fibre reinforced laminar composites
Christian M. Laurent1, Colin Palmer2, Richard P. Boardman3, Gareth Dyke1,4 and Richard B. Cook5
Accepted: 30 September 2014
Search Result:
however only small grids of indents were performed and did not span across multiple laminae t bachmann 2014 personal communication
the aim of our study was to determine whether nano indentation can indeed be used to identify laminae within feather rachises and to investigate differences in the laminae of three primary feathers from a mute swan cygnus olor a bald eagle haliaeetus leucocephalus and a partridge perdix perdix
ultra high resolution computed x ray tomography ct has also become an increasingly useful tool in materials science

ultra high resolution computed x ray tomography ct has also become an increasingly useful tool in materials science
as resolution capabilities become ever smaller it has enabled nano scale imaging of individual fibres in composite materials and the use of appropriate software to measure laminar proportions and fibre orientation which in this case will validate the nano indentation method

material and methods 21

the new longitudinal sections surface was also polished to a 1 mm diamond finish

nano indentation nano indentation testing was performed on a micromaterials vantage nano indentation system capable of resolving force and displacement to 3 nn and 0001 nm respectively the indentations were run in depth control to 500 nm at a loading rate of 005 mns 1 with a 40 s hold at peak load to reduce the effect of viscoelasticity on the modulus results

the indentations were spaced 15 microns apart to avoid overlapping interaction volumes with the number of indents varying between feathers to ensure the map spanned the width of the rachis cross section

the indentations were spaced 15 microns apart to avoid overlapping interaction volumes with the number of indents varying between feathers to ensure the map spanned the width of the rachis cross section

the diamond area function of the berkovich diamond tip was calibrated before the first experiment and the indenter was checked on a fused silica standard before each new map

a single map was performed across each feather rachis with two additional maps on c

using the 20x objective the achieved resolution was 794 nm per voxel

results the er results from indentation maps of the second c

olor calamus cross section and corresponding longitudinal section are shown in table 1

the er results from the outer and inner laminae were reversed when tested in the two different orientations

figure 3 shows cross sectional indentation maps of the three feathers

three laminae were detected in p

curves which conform to this power law are well suited to evaluation by the method in oliver pharr their method uses the tangential slope dp/dh of the unloading phase of the load/ displacement line to compute the reduced modulus where a is the contact area and s is the measured stiffness dp/dh

reduced modulus er is used in place of youngs modulus because calculating a reduced modulus accounts for the compliance of a non rigid indenter the indenter was calibrated on fused silica before each array of indentations

high resolution ct of a matchstick sample of c

high resolution ct of a matchstick sample of c

olor confirms fibre orientation anisotropy identified by the nano indentation

figure 5 shows three clear layers within the rachis cross section represented graphically in figure 6

the fibres of the innermost layer are directed between 5 and 10 relative to the middle layer and also run longitudinally

discussion our study is the first to use nano indentation mapping to identify laminae in bird feather rachises

this technique however relies on the differing orientations of the b keratin fibres within the laminae identified by lingham soliar et al and their resulting anisotropy for differentiation

our findings thus corroborate previous works which have demonstrated that laminae contain differentially orientated b keratin fibres in an amorphous protein substantia

in all five datasets we have produced there is deviation around mean er for a particular line of indentations across a rachis section resulting from inhomogeneity within the matrix of the laminae

the 500 nm indentations interrogate an interaction volume equivalent to a half sphere with a 5 mm radius

three rachis laminae four including an outer lipid membrane have previously been reported in a peacock feather using x ray microdiffraction busson et al observed two feather laminae with a thin 25 mm amorphous protein lamina in between

figure 2ab does arguably show a thin lamina at the interface although it is more likely that none of the indentations landed directly on this proposed amorphous lamina and this is an artefact of the graphing softwares interpolation algorithm
the scanning electron microscope sem image presented by lingham soliar after microbial degradation does show busson et als two feather laminae but the nature of the degradation combined with their imaging method make it difficult to distinguish the amorphous lamina from the embedding substantia

the scanning electron microscope sem image presented by lingham soliar after microbial degradation does show busson et als two feather laminae but the nature of the degradation combined with their imaging method make it difficult to distinguish the amorphous lamina from the embedding substantia
it should be noted that we do not observe an amorphous lamina in between fibre laminae on any nano indentation maps nor in the ct scan all of which show an abrupt transition from one fibrous lamina to another
if a third feather lamina was missed by this approach there are a number of possibilities it may be difficult to differentiate laminae if the matrix degradation leaves fibres loose and able to change their orientation slightly or at least enough to hide subtle variations in orientation

if a third feather lamina was missed by this approach there are a number of possibilities it may be difficult to differentiate laminae if the matrix degradation leaves fibres loose and able to change their orientation slightly or at least enough to hide subtle variations in orientation
alternatively if differences in orientation are subtle to begin with they might not be sufficiently different enough to diffract busson et als x rays or even show up on our modulus mapping if the two innermost laminae are at similar orientations to the indenter tip despite being relatively different they may also have been reported as a single lamina in the second indentation of c
further work is clearly needed to compare and contrast the two approaches as neither lingham soliar et al nor busson et al reported the location in the wing from where they sampled feathers

Article Info:
i6 The structure and mechanics of Moso bamboo material
P. G. Dixon and L. J. Gibson
Accepted: 30 June 2014
Search Result:
the volume fraction of solid in the vascular bundles also increases radially towards the outer part of the culm as the vessels become smaller both the increasing volume fraction of vascular bundles and the increasing volume fraction of solid within each vascular bundle lead to a pronounced radial density gradient in the bamboo culm with denser tissue towards the outer part of the culm
the outer epidermal layer of bamboo is rather hard with a waxy surface this radial density gradient has a profound effect on the mechanical properties of bamboo the axial tensile youngs modulus varies from about 5 to 25 gpa and the axial tensile strength varies from about 100 to 800 mpa for specimens taken from the inner and outer culm respectively the compressive and flexural properties of bamboo along the grain have also been investigated the compressive strength increases with the height on the culm and decreasing moisture content for example lo et al find the compressive strength of moso bamboo increases with height along the culm from 45 to 65 mpa and lee et al find the flexural strength increases from 70 mpa in the green state to 103 mpa for air dried bamboo most of these investigators probe the mechanical variation with height and/or some other variable age presence of nodes and moisture content
however very few studies attempt to capture the radial variation of mechanical properties other than tensile properties in the axial direction

in this study we characterize the microstructure of moso bamboo quantitatively and measure the axial youngs modulus in bending or modulus of elasticity e axial modulus of rupture in bending axial

compressive strength radial compressive strength and tangential compressive strength to assess the effect of the radial density gradient on the mechanical properties ofmoso bamboo
in addition nanoindentation experiments were performed to determine whether there is a gradient in modulus or hardness at the cell wall level
finally models relating the structure and mechanical properties were developed and compared with the experimental data

r/a 025 050 and 075 respectively
oliver pharr analysis of the unloading curve was used for the determination of reduced moduli the raw data suggested that the reduced modulus and hardness showed little positional effects in either the radial or longitudinal direction
indents were filtered for outliers using the interquartile range iqr of the effective depth computed from all the indents

we note that the higher magnification images in figure 13 clearly show the vascular bundles collapsing with little deformation visible in the parenchyma tissue
the average reduced modulus and hardness from nanoindentation of the solid cell wall in the sclerenchyma fibres are 14923 gpa and 28964 mpa table 1 respectively similar to literature values of 160315 gpa and 360 104 mpa for moso bamboo values given as averagesd
while there are slightly different values with varying radial and longitudinal positions all values of both reduced modulus and hardness were within 1 sd

the average reduced modulus and hardness from nanoindentation of the solid cell wall in the sclerenchyma fibres are 14923 gpa and 28964 mpa table 1 respectively similar to literature values of 160315 gpa and 360 104 mpa for moso bamboo values given as averagesd
while there are slightly different values with varying radial and longitudinal positions all values of both reduced modulus and hardness were within 1 sd
of the mean suggesting that the sclerenchyma fibre properties are roughly constant within the culm

the mechanical properties of moso bamboo can be compared to north american softwoods commonly used for structural purposes table 2 such as eastern white pine douglas fir and white spruce
youngs modulus is comparable except for douglas fir which is higher than moso bamboo while the modulus of rupture and compressive strength of the bamboo are much higher than the above softwoods the average density of moso bamboo is significantly higher than all of these softwoods comparison of transverse data is unclear as most transverse data on wood is performed using a practical engineering test that has some indentation character it is important to note that the average density of the moso bamboo studied is significantly higher than that of any of these softwoods and is similar to that of northern red oak a commercially important north american hardwood
our data for moso bamboo indicates that it is 31% stronger in bending 15% less stiff and 48% stronger in compression than red oak it is worth noting that while the compressive strength results are impressive youngs modulus determined from bending tests is comparatively low when compared with wood suggesting that structural moso bamboo members in compression could be limited by buckling

Article Info:
i8 2014 Housing tubes from the marine worm Chaetopterus sp.: biomaterials with exceptionally broad thermomechanical properties
Shah DU, Vollrath F, Porter D, Stires J, Deheyn DD
Accepted: 19 June 2014
Search Result:

the complex structureproperty relations observed indicated that the worm has evolved to produce a tubular housing structure which can i function stably over a broad range of temperatures ii endure mechanical stresses from specific planes/axes and iii facilitate rapid growth or repair
introduction the marine worm chaetopterus variopedatus polychaete chaetopteridae lives in self constructed cylindrical u shaped parchment like tubes typically buried in soft seafloor sediment with both the tapered extremities exposed
regarded as one of the most structurally specialized of all annelids chaetopterus animals have a cosmopolitan distribution typically occurring in mud or sand flats in the intertidal zone or in shallow coastal waters less than 50 m deep notably dense aggregations of similar chaetopterids and their characteristic tubes have been found at even deep sea sites greater than 1000 m deep near hydrothermal vents around mud volcanoes at cold seep sites and around whale falls this suggests that not only the chaetopterid worm but also their self secreted tubes have a broad range of environmental tolerance including high subsea pressure temperature salinity and ph levels our own anecdotal observations of tubes kept in aquaria of solutions of seawater 70% ethanol or various ph for several years also draw attention to the sturdy nature of the tubes as these show no apparent change in shape size texture among other properties

Article Info:
i9 Compliant threads maximize spider silk connection strength and toughness
Avery Meyer1, Nicola M. Pugno2,3 and Steven W. Cranford1
Accepted: 16 June 2014
Search Result:
in webs as in engineered structures brittle like small deformation failure should be avoided
finally with the addition of pre stretch we note a distinct transition from a nonlinear stiffening response to a more softening like response
the hyperelastic character of the global response changes from an extreme stiffening like behaviour at low pre strains most similar to the constitutive law figure 2 to a near constant stiffness with sudden yield at high pre strains

stiffening behaviour in the anchorage is desired owing to the relative initial compliance
as the material transitions from linear elastic to softening behaviour there is little change in the connection capacity for the strain regime explored
this can be attributed relative stiffness of a plastic softening like behaviour at small strains for the case here where 1ult 2 at 1 015 the ratio of anchor thread tangent stiffness with b 05 to adhered thread tangent stiffness with b 3 is greater than 100 see methods section

as the material transitions from linear elastic to softening behaviour there is little change in the connection capacity for the strain regime explored
this can be attributed relative stiffness of a plastic softening like behaviour at small strains for the case here where 1ult 2 at 1 015 the ratio of anchor thread tangent stiffness with b 05 to adhered thread tangent stiffness with b 3 is greater than 100 see methods section
in effect using the constitutive law expressed by equation 24 the anchor thread approaches the rigid condition for b 1 at small strains

this is a key finding in consideration of the geometry of the web wherein silk connections are typically perpendicular viscid and dragline connections for example and the variation/ control in delamination angle greatly influences joint strength
we have further explored the intrinsic optimization mechanism of potential attachment structures using an extension of the theory of multiple peeling for soft substrates

in comparison with attachment to rigid substrates a compliant anchor thread increases the work to delaminate and behaves like a surface with increased adhesion strength captured by the effective peeling angle a a and modified parameter l independent of the properties of the adhered thread

the simple anchorage is colarchic two branched symmetrical and homogeneously adhesive across a similar compliant thread
it is an adhesive anchorage because it allows a force f to be transmitted to a solid substrate through adhesive forces at the material interface eg no penetration of material entanglement symmetrical because the initial angles a on both sides are equal and it is colarchic because it has no hierarchy eg a single connection
the model represents the most basic geometry of splayed silk connections that engage adhesive forces at a silk thread interface

constitutive model we consider the general stressstrain relation expressed by equation 24 defined by the ultimate stress sult strain 1ult a hyperelastic parameter b and a scaling factor k
for stiffening b 1 whereas softening occurs for b 1 and linear elastic behaviour when b 1
because we are interested in forces we simply multiply by cross sectional area and rearrange to separate the constants

Article Info:
i10 Lee N, Horstemeyer MF, Rhee H, Nabors B, Liao J, Williams LN. 2014 Hierarchical multiscale structure–property relationships of the red-bellied woodpecker (Melanerpes carolinus) beak. J. R. Soc. Interface
11: 20140274. http://dx.doi.org/10.1098/rsif.2014.0274
Search Result:
at the nanoscale a wavy gap between the keratin scales similar to a suture line was evidenced in the rhamphotheca the middle foam layer joins two dissimilar materials and mineralized collagen fibres were revealed in the inner bony layer
the nano and micro indentation tests revealed that the hardness associated with the strength modulus and stiffness of the rhamphotheca layer approx
470 mpa for nano and approx

also the elastic moduli of the woodpecker toucan hornbill and java finch are represented in table 1
the elastic moduli of the beaks of woodpeckers toucans and hornbills were obtained from nanoindentation testing while that of java finch was obtained from a double indentation technique
the different mechanical properties of the birds reveal different functions and uses of each beak

the porosity and area fraction in the beak were measured via analysing two dimensional images using image j software national institutes of health bethesda md usa
microindentation and nanoindentation tests were conducted on the lower beak in order to evaluate the mechanical properties
a vickers hardness tester leco corporation st joseph mi usa with a pyramidal diamond tip was used to examine the beak microhardness

microindentation and nanoindentation tests were conducted on the lower beak in order to evaluate the mechanical properties
a vickers hardness tester leco corporation st joseph mi usa with a pyramidal diamond tip was used to examine the beak microhardness
the applied maximum load was 100 gram force

this procedure was employed due to the creep behaviour of viscoelastic materials

the hardness was defined by the following equation where p is the maximum applied load n and a the resultant projected contact area
for a berkovich tip a is calculated from the ideal tip area function 254h2 and h is the maximum displacement

for a berkovich tip a is calculated from the ideal tip area function 254h2 and h is the maximum displacement
the reduced elastic modulus was derived from the initial unloading contact stiffness where the reduced modulus er is derived from the displacement from both the specimen and the indenter
the reduced elastic modulus is given by where e and n are the elastic modulus and poissons ratio of the specimen and the indenter respectively

the reduced elastic modulus was derived from the initial unloading contact stiffness where the reduced modulus er is derived from the displacement from both the specimen and the indenter
the reduced elastic modulus is given by where e and n are the elastic modulus and poissons ratio of the specimen and the indenter respectively
structure of the woodpeckers beaks 311

from the image of figure 5c the length the d period in the fibres was measured to be in the range of 6070 nm
the measured period falls into the range of collagen fibril d period although it is not the exact 6769 nm band that is often reported in soft connective tissues the stereological effect might explain this variation of d period estimation eg a slight tilt can change the period estimation
the morphology of the banded fibres under sem and the periodic length all point to the likelihood that the fibres contained in the foam layer are collagen fibres

microindentation tests were performed to garner micromechanical properties
indentation tests were conducted only on the rhamphotheca and bony layers because the area of the foam layer was not large enough to conduct the tests
the average value of the microhardness was 032001 gpa for the rhamphotheca layer and 064007 gpa for the bony layer table 1

to examine the gradient of the nanomechanical properties experiments were performed on four different beak locations from the tip to the root and the results are depicted in figure 8
the results show a decrease in the hardness as the location changes from the tip to the root of the beak for the rhamphotheca
the bony layer did not show a change with respect to location

the nanohardness was measured as 040008 gpa for the rhamphotheca 024014 gpa for the foam layer and 116019 gpa for the bony layer table 1
the relatively high mineral content of the bony layer was responsible for it being three times harder than the keratin part rhamphotheca
the average measured values of reduced elastic moduli were 8711 gpa at the rhamphotheca 6525 gpa at the foam layer and 30236 gpa at the bony layer table 1

macrostructuremechanical property relationships the woodpeckers beaks are functionally graded materials through the length of the beak from three perspectives i the geometry of the cross section of the beak changes as shown in figure 2a ii the area fractions of the rhamphotheca foam and bony portion change as shown in figure 2b and iii the associated mechanical properties of the rhamphotheca foam and bony portion are different
according to the rule of mixtures which was developed for composite materials and can be applied for layered materials one can estimate the modulus and hardness at each point along the beak with the

correlating area fractions and strength/hardness of each sample and where a represents the area fraction e represents elastic modulus and h is the hardness value
the total strength of the composite increased from the beak tip to the root

micromechanically the microhardness of the rhamphotheca of the woodpecker is about 50% greater than those of other birds such as the toucan hornbill and starling
it is reported that dark coloured beaks have a greater hardness than light coloured beaks so the dark beaks are less susceptible to wear as the colour of woodpeckers beaks is predominantly black then one would expect a greater hardness if the european starlings study is consistent with woodpeckers beaks
our study also shows that the microhardness of the core part of the woodpeckers beaks is indeed two to three times greater than those of the toucan and hornbill beaks table 1

our study also shows that the microhardness of the core part of the woodpeckers beaks is indeed two to three times greater than those of the toucan and hornbill beaks table 1
clearly in terms of the structurefunction relationship hardness differences would be expected in the woodpeckers beaks versus the toucan and hornbill because the high rate shocks would be much greater for the woodpecker
nanostructuremechanical property relationships from the nanostructure of the woodpeckers beaks as shown in figure 4e one can observe that mechanical anisotropy arises because the fibres are oriented parallel to the pecking direction

various values were measured across the span of the bony layer
the hardness and reduced elastic modulus were greater at the location where there was mostly a mineral matrix and they were lower where there was a greater fibre density but lower mineral content
comparison to other birds beaks the structure of the woodpeckers beak is different from other birds beaks such as the chicken and toucan

the woodpecker beak is a structural biocomposite having three layers rhamphotheca outer keratin shell middle foam layer and inner bony layer
along the beak from posterior to anterior the area fraction of these three layers gradually changes so the aggregate modulus and aggregate hardness are gradients
the rhamphotheca is made up of elongated keratin scales and the microhardness of the rhamphotheca was measured to be approximately 323 mpa and the nanohardness approximately 470 mpa

Article Info:
i11 Rafsanjani A, Stiefel M, Jefimovs K, Mokso R, Derome D, Carmeliet J. 2014 Hygroscopic swelling and shrinkage of latewood cell wall micropillars reveal ultrastructural anisotropy. J. R. Soc. Interface 11: 20140126.
http://dx.doi.org/10.1098/rsif.2014.0126
Search Result:
particularly in the transverse plane of wood the main moisture induced deformations occur in the s2 layer sorption of water molecules in the hydrophilic matrix pushes the constituents apart resulting in formation of new pores and giving rise to swelling and softening of the matrix however the s2 material is not available for direct measurement since its microscopic dimensions make isolation by manual manipulation a challenge
alternatively the focused ion beam fib technique allows micro scale samples to be prepared from individual structural components of the wood cell wall which otherwise are not accessible for testing

by transforming the swelling coefficients tensor with a rotation r equal to the mfa around the axis we can determine b and b from the components of the transformed swelling coefficient tensor which is written in matrix form as

the preferential direction of stiff cellulose microfibrils within the s2 layer is expected to restrain swelling probably more in parallel direction than normal to the cellwall thicknessconsidering that the s2 layer is composed of an almost equal quantity of stiff non swelling crystalline cellulose ef 134 gpa and a soft polymeric matrix em 2 gpa for mfa 30 the resulting anisotropy in the transverse swelling properties of the s2 layer is approximately according to equations 41 and 42 b/b 13
therefore considering the small mfa of s2 layer which is generally in the range 1030 we may conclude that the mfa would thus be a factor of minor importance for swelling/shrinkage anisotropy in the transverse direction

Article Info:
i13 De Tommasi D, Millardi N, Puglisi G, Saccomandi G. 2013 An energetic model for macromolecules unfolding in stretching experiments. J R Soc Interface 10: 20130651.
http://dx.doi.org/10.1098/rsif.2013.0651
Search Result:
introduction the past decade has shown a significant theoretical and experimental effort in the analysis of the thermomechanical behaviour of macromolecular materials such as muscle tissues spider silks polymers and biopolymers polysoaps and silks in general a common property of these materials is that their macroscopic history dependent dissipative response is the result of complex evolutions of semicrystalline microstructures
these are constituted by flexible polymeric or protein macromolecules reinforced by strong and stiff crystals eg in the form of fillers or b sheets undergoing typical hardsoft transitions owing to the unravelling of hard crystal domains into soft unfolded entropic domains
the deduction of predictive models describing the mechanical behaviour of macromolecular materials connecting the macroscale response with the mesoscale behaviour is crucial not only for the description of the fundamental polymeric and biopolymeric existing materials but also in the perspective of the design of new bioinspired or reconstructed biological materials

these transitions have a main role in the energetics of the chain unfoldings owing to the two following effects first there is the enthalpic contribution of the transition itself debonding energy then there is an entropic contribution associated with the variation of the microstructure leading to increased contour lengths of the chains
in the final stage leading to the material failure the behaviour is regulated by the entropic hardening of the fully unfolded macromolecules primary structure
to fix the ideas in figure 1 we schematically show a typical afm single molecule stretching experiment reproduced from on an engineering reconstructed macromolecule

the macromolecule follows elastically the first equilibrium path oa in figure 1 with the external workwaccumulated as elastic energy fe dw dfe
at the first b sheet unfolding path obhere approximated as instantaneous with no external work wthere is an internal energy discontinuity area oab that by energy conservation equals the unfolding energy q1 of the first hard domain
similar considerations can be extended to the next elastic and dissipative steps so that the total dissipation is based on previous considerations and inspired by the model in where the authors describe the hysteresis of filled polymers and spider silks here we propose an energetic approach and describe the unfolding macromolecule as a two state material similarly a two state energetic approach was proposed in to describe the helix coil transition of polypeptide chains regulating the damage of multi block copolymers

in this work we obtained an analytic solution in the thermodynamic limit of a large number of breakable links

recently a statistical mechanics approach for a discrete two well finite size chain has been proposed describing the forceextension diagrams of polymeric chains under assigned force soft device or end to end length hard device
in particular numerical results let us describe the sawtooth behaviour observed in a hard device

recently a statistical mechanics approach for a discrete two well finite size chain has been proposed describing the forceextension diagrams of polymeric chains under assigned force soft device or end to end length hard device
in particular numerical results let us describe the sawtooth behaviour observed in a hard device
in the field of protein mechanics on the basis of the approach proposed in an important step in the comprehension of the energy competition between the unfolding and entropic energy terms has been delivered in in this paper the author models the unfolding of a biomacromolecule as a chain composed of folded and unfolded domains both elastic with gaussian type response combined with an ising type unfolding energy

it is important to point out that such experiments often show rate dependent effects because at the afm loading time scale the b sheet unfolding transitions represent out of equilibrium events eg the analysis performed by extending the kramer reaction theory in or the md analysis in here aimed again to the deduction of a fully analytical approach following we take care of the described rate dependence by considering effective rate dependent unfolding energies
energetic assumptions because in macromolecules the hard crystal unfolding is typically an all or none transition as confirmed also from the size of periodicity of the experimental unfolding lengths in the case of titin we model the molecule as a discrete lattice of n two state rigid folded/entropic unfolded links see the scheme in figure 1
the folded/unfolded state of the chain is assigned by a set of internal variables xi i 1 n such that xi 0 xi 1 denotes the folded unfolded state of the ith domain

thus using 21 and 22 the total elastic energy is and correspondingly the total forcedeformation relation is finally according to 23 the total energy is we follow a griffith like approach minimizing the total unfolding fracture energy plus elastic entropic energy and based on 11 we assume that the observed solutions are the global minima of ftot in 36
the first important step is to justify the experimental observation that the hard crystals unfold one at a time resulting in a constant increase of the contour length see to obtain this result we begin by evaluating the solution of the equations and get the intersection lengths l
the searched result follows by the observation that so that if nu is the branch corresponding to the global minimum at assigned end to end length l by increasing l it loses its global stability at the intersection with the equilibrium branch nu 1 figure 3

the searched result follows by the observation that so that if nu is the branch corresponding to the global minimum at assigned end to end length l by increasing l it loses its global stability at the intersection with the equilibrium branch nu 1 figure 3
thus the chain unfolds with a sequence of single hard domain transitions at the threshold assigned by in this formula and in the following we omit the nu dependence of lc
in 37 we introduced the main non dimensional parameter of the model representing a measure of the ratio between the elastic and fracture energy of the single b sheet

by increasing further the assigned length the system follows the new branch until another sudden transition to the branch nu 2 is observed when it becomes energetically favourable path bb
similar transitions with single domain unfoldings are then observed until all the crystals unfold and the system shows a hardening behaviour owing to the entropic elasticity of the fully unfolded chain curve fg in figure 4
regarding the behaviour of the system under unloading because typically at the afm loading rate no refolding is observed we assume irreversible hardsoft transitions

the memory of the system is then restricted to the only maximum value attained in the past by the end to end length l

finally we observe that the theoretical model shows a softening behaviour during the unfolding regime in the sense that the unfolding force decreases with the number nu of unfolded domains

this can be analytically proved by observing that in view of 310 we have nevertheless the experimental behaviour of stretch induced unfolding of macromolecules shows a variable behaviour with typically nearly constant unfolding forces but with possibly increasing decreasing or non monotonic transition thresholds see for titin macromolecule and for artificial elastomeric protein

this can be analytically proved by observing that in view of 310 we have nevertheless the experimental behaviour of stretch induced unfolding of macromolecules shows a variable behaviour with typically nearly constant unfolding forces but with possibly increasing decreasing or non monotonic transition thresholds see for titin macromolecule and for artificial elastomeric protein

in the following section we discuss this issue and propose an extension of our model able to reproduce the hardening effect observed eg in titin macromolecule unfolding

unfolding energy hierarchy the described variable observations on subsequent transition forces force plateaux hardening softening have been given different physical interpretations

in the following section we discuss this issue and propose an extension of our model able to reproduce the hardening effect observed eg in titin macromolecule unfolding

unfolding energy hierarchy the described variable observations on subsequent transition forces force plateaux hardening softening have been given different physical interpretations

a hierarchy of unfolding energies of the unfolding crystals may be simply due to inhomogeneity effects of the crystal domains showing variable bond breaking barriers possibly owing to interfacial energy effects another important effect is anisotropicity of the crystals with respect to the force directiondifferent paths in the wiggly energy landscape lead to different unfolding energy barriers so that different unfolding forces can be induced by variable crystal orientations in the macromolecule

a hierarchy of unfolding energies of the unfolding crystals may be simply due to inhomogeneity effects of the crystal domains showing variable bond breaking barriers possibly owing to interfacial energy effects another important effect is anisotropicity of the crystals with respect to the force directiondifferent paths in the wiggly energy landscape lead to different unfolding energy barriers so that different unfolding forces can be induced by variable crystal orientations in the macromolecule

another known effect that can induce hardening is the so called n effect see that based on statistical considerations leading to an unfolding force growth owing to a progressively reduced number of folded crystals available for unfolding in the macromolecule for growing elongations

to check the possibility of variable unfolding energies following the energetic decomposition schematized in figure 1 based on the experimental forcedisplacement unfolding diagrams we may estimate the fracture energy of each unfolding event using the relation where represents the variable fracture energy of the nu th folded domain and hu lunu/lcnu is the corresponding unfolding strain

it is important to remark that in the case of increasing unfolding energies the di block approximation can fail with the order of unfolding that is regulated by the competition of mixing energy interfacial energy effects and variability of the unfolding energy

this competition regulates in the quasi static regime the order of transition and the hardening softening or non monotonic law of the successive unfolding force

aimed again at a fully analytical result we here suppose that both the mixing entropy contribution and the interfacial energy term are negligible when compared with the unfolding energy increment dq so that the unfolding evolution strategy is regulated by the unfolding energy hierarchy and we may again describe the stretching behaviour of the macromolecule by minimizing an energy as in 36 but with variable unfolding energy increments

Article Info:
i14 Checa AG, Bonarski JT, Willinger MG, Faryna M, Berent K, Kania B, Gonza´lez-Segura A, Pina CM, Pospiech J, Morawiec A. 2013 Crystallographic orientation inhomogeneity and crystal splitting in biogenic calcite. J R Soc Interface 10: 20130425.
http://dx.doi.org/10.1098/rsif.2013.0425
Search Result:
phase images reveal a clear contrast between the nanounits and the covering membrane with the former appearing much brighter
such a contrast indicates that nanounits are harder than the membrane
all together these observations indicate that the membranes most likely have an organic nature

Article Info:
i15 Shape optimization in exoskeletons and endoskeletons: a biomechanics analysis
David Taylor* and Jan-Henning Dirks
Published online 12 September 2012
Search Result:
the elliptical shape of the locust tibia certainly has a positive effect as noted earlier
a further limitation is that we considered arthropod bones to be made from a single layer of cuticle when in fact they consist of several layersin particular a relatively hard stiff exocuticle and a much softer endocuticle this layering can affect other aspects of the mechanics of cuticle structures but it is probably not a serious limitation in our case because the analysis is self consistent we used material property values that also assumed a single layer
it would be very interesting to predict the optimum values for the thicknesses of these different layers though at present there is not enough experimental data to make the attempt worthwhile

Article Info:
i17 Epidermis architecture and material properties of the skin of four snake species
Marie-Christin G. Klein* and Stanislav N. Gorb
Published online 15 August 2012
Search Result:
on the basis of structural and experimental data it was previously demonstrated that the snake integument consists of a hard robust inflexible outer surface oberhautchen and b layer and softer flexible inner layers a layers
it is not clear whether this phenomenon is a general adaptation of snakes to limbless locomotion or only to specific conditions such as habitat and locomotion

even though snakes from different lineages have different body size scale dimension body shape and body mass previous morphological examinations have shown that their epidermis basically consists of six layers overlying the dermis the following epidermis layers from the outer scale surface towards the dermis are known figure 1c i the oberhautchen ii b layer iii mesos layer iv a layer v lacunar tissue and vi clear layer on the basis of structural data it has been previously assumed that the epidermis of snakes consists of a hard robust inflexible outer surface oberhautchen and b layer and soft flexible inner layers a layers in our previous study we used the nanoindentation technique which has been previously used to evaluate local mechanical properties of various biological materials both the effective elastic modulus eem and hardness hd of the outer and inner epidermis layers of the ventral scales of the kenyan sand boa gongylophis colubrinus exuvium were characterized the results obtained provided evidence for the presence of a gradient in material properties in the g although it is known that the epidermis of snakes consists of six main layers there has not been a single study comparing the cross section architecture of these layers between different snake species inhabiting different environments and preferably using different modes of locomotion

it is suggested that both the spacing and dimensions of these structures play an important role in the ability to detect and enhance the absorption of infrared radiation in sliding contacts local stress concentrations on the surface lead to material fatigue and failure
the surface is more effective against abrasion wear if it has a gradient in material structure and properties because it leads to a more uniform stress distribution and thus to the minimization of local stress concentrations a hard inflexible surface material easily forms cracks under pressure whereas a soft flexible system will be easily worn off under shear stress
as we have previously assumed the gradient in material properties from a stiff surface to a soft depth will improve wear resistance in snake skin by combining the advantages of stiffness and flexibility

the surface is more effective against abrasion wear if it has a gradient in material structure and properties because it leads to a more uniform stress distribution and thus to the minimization of local stress concentrations a hard inflexible surface material easily forms cracks under pressure whereas a soft flexible system will be easily worn off under shear stress
as we have previously assumed the gradient in material properties from a stiff surface to a soft depth will improve wear resistance in snake skin by combining the advantages of stiffness and flexibility
the exuvium surface geometry of shed epidermis does not differ from that of a live animal m

morelia viridis was reared in a 210x220x130 cm terrarium with plants and branches as the substrate nanoindentation the experimental set up was the same as described in klein et al in which the nano indenter sa2 mts nano instruments oak ridge tn usa equipped with the continuous stiffness measurement csm technique allowing testing of soft biological tissues or gels was used for dynamic indentation tests nanoindentation is a technique to measure the mechanical properties of small volumes of material from obtained forcedisplacement curves both hd and elasticity modulus of materials can be determined at penetration depths ranging from several hundred nanometres to few dozens of micrometres hd and elasticity were acquired from the following equations where h is hardness pmax is the maximal load and ac is the contact area with eeff being the reduced or effective elastic modulus b the correction factor for the indenter form and s the contact stiffness

nanoindentation the experimental set up was the same as described in klein et al in which the nano indenter sa2 mts nano instruments oak ridge tn usa equipped with the continuous stiffness measurement csm technique allowing testing of soft biological tissues or gels was used for dynamic indentation tests nanoindentation is a technique to measure the mechanical properties of small volumes of material from obtained forcedisplacement curves both hd and elasticity modulus of materials can be determined at penetration depths ranging from several hundred nanometres to few dozens of micrometres hd and elasticity were acquired from the following equations where h is hardness pmax is the maximal load and ac is the contact area with eeff being the reduced or effective elastic modulus b the correction factor for the indenter form and s the contact stiffness
owing to the variation of material structure in the depth of biological samples mechanical properties might fluctuate considerably with depth therefore the csm method was applied which allows measurements of mechanical properties as a function of depth in our previous work we successfully applied this method for mechanical characterization of biological samples such as insect cuticle plant cuticle and snake skin 23

cenchria and m
viridis were cut from ventral mid body regions attached to an aluminium sample holder as described in klein et al in one of the two ways i the osl oberhautchen/b layer of the oss facing the indenter tip ii the isl clear layer of the oss facing the tip
this way the mechanical properties into the direction perpendicular to the surface and to the keratin fibres were obtained

a total of 2550 individual indentations on the posterior margin of the scale were carried out on each individual scale
the samples were loaded under constant strain rate conditions with a rate of 002 s 1 to a maximum penetration depth of 10 mm for further details see to ensure proper surface detection the procedure described in deuschle et al was used
scanning electron microscopy of cross sections ventral scales of g

differences between the a layer and the lacunar tissue are difficult to make out and the clear layer is not identifiable
nanoindentation measurements different penetration depths were used for different species in the nanoindentation tests owing to different thickness of their exuvia
indentation data from the first 710% of the total sample thickness can only be used for analysis in order to avoid the influence of the material properties of the sample holder

nanoindentation measurements different penetration depths were used for different species in the nanoindentation tests owing to different thickness of their exuvia
indentation data from the first 710% of the total sample thickness can only be used for analysis in order to avoid the influence of the material properties of the sample holder
therefore data were used for penetration depths of 3001800 nm for g

indentation data from the first 710% of the total sample thickness can only be used for analysis in order to avoid the influence of the material properties of the sample holder
therefore data were used for penetration depths of 3001800 nm for g
colubrinus 3001200 nm for l

cenchria and 3001500 nm for m
the results table 2 show that the osl is significantly harder and has a higher elasticity modulus when compared with the isl p 0001 mannwhitney rank sum test t test in all four species studied
in order to compare material properties of the inner and outer layers between species data were selected from the same indentation depths 300700 nm

the results table 2 show that the osl is significantly harder and has a higher elasticity modulus when compared with the isl p 0001 mannwhitney rank sum test t test in all four species studied
in order to compare material properties of the inner and outer layers between species data were selected from the same indentation depths 300700 nm
table 3 shows the results of this analysis

californiae and e
when comparing the values of the selected indentation depths as listed earlier the results between species are also significantly different p 0001 bonferroni t test with a few exceptions
as mentioned already the modulus values of the isl of l

viridis 300400 10001100 and 15001600 nm table 7
this comparison revealed a clear trend from soft and flexible inner layers to stiffer and less elastic outer layers for all four species
colubrinus table 4 there was no significant difference for the hd and modulus values between the three intervals for indentation from the inside of the scale

this comparison revealed a clear trend from soft and flexible inner layers to stiffer and less elastic outer layers for all four species
colubrinus table 4 there was no significant difference for the hd and modulus values between the three intervals for indentation from the inside of the scale

this would suggest a smooth transition between layers

cenchria table 6 were not significantly different except for the hd values for the osls between depth intervals of 400500 and 13001400 nm
it is possible that the indentation depth selected for this species was not sufficient to reveal differences between different layers of the exuvium
the modulus and hd values for the osls in m

viridis table 7 were significantly different except for the hd results between the intervals 10001100 and 15001600 nm
there was a clear trend from soft and flexible inside layers to hard and inflexible outside ones
the modulus values for the isls exhibit a significant difference between the outer and inner most intervals 300400 and 15001600 nm and not between the others

the modulus values for the isls exhibit a significant difference between the outer and inner most intervals 300400 and 15001600 nm and not between the others
this implies a smooth transition from the soft to the harder regions
the hd values on the other hand for the inner scale values are significantly different except for the intervals 10001100 and 15001600 nm which implies that the interval 300400 nm is very soft in comparison with the others and that the transition is not very smooth

this implies a smooth transition from the soft to the harder regions
the hd values on the other hand for the inner scale values are significantly different except for the intervals 10001100 and 15001600 nm which implies that the interval 300400 nm is very soft in comparison with the others and that the transition is not very smooth
discussion our morphological results and nanoindenation data obtained on the ventral scales provide strong evidence for the existence of a gradient in material properties in the four snake species studied

these have a temporarily high synthesis rate within the oberhautchen and b layer which decreases in the underlying layers the hd and modulus values were relatively high close to the surface
because the tip oscillation independent of the amplitude affects the contact formation as well as the contact stabilization for displacements between 0 and approximately 300 nm indentation depth the results obtained at these shallow depths were not reliable to make any conclusion previous authors assumed that material properties are important for abrasion resistance it was previously shown that a and b keratin differ not only in their distribution within the snake epidermis but also in their chemical structure and therefore presumably also in material properties differences in the chemical structure indicate that the b keratin molecule may be more stiff and inelastic when compared with a keratin
that is why it is plausible to assume that materials containing a higher proportion of b keratin are more robust and therefore posses a stronger resistance against damage

later using nanoindentation we showed that both the eem and hd of the osls which are made of b keratin is higher compared with the isls which have a keratin integrated
it was not only experimentally demonstrated that materials containing b keratin is indeed harder and less stiff than materials containing a keratin but that there is a gradient in material properties of the integument from a hard and inflexible outside to a soft and elastic inside 41
material property gradients in other biological tissues wang weiner as well as fong et al showed that in the mineralized tissue of the mammalian tooth layered organization with the gradient of material properties is a key mechanism of the wear resistance

material property gradients in other biological tissues wang weiner as well as fong et al showed that in the mineralized tissue of the mammalian tooth layered organization with the gradient of material properties is a key mechanism of the wear resistance

the visible part of the tooth above the gum is enveloped by the enamel which is hard and highly resistant to wear its primary function is to protect the underlying dentin the tooths major component and the more flexible soft pulp using indentation tests on pulp and pulpless dog teeth with the knoop indentor fusayama maeda showed that the dentin hd in pulpless teeth is lower than that of vital teeth
the authors assume this to be due to the absence of the pulp functioning as a damping layer for the stiffer dentin bruet et al used instrumented nanoindenation to measure the multilayered structure of four different organicinorganic nanocomposite material layers of the scales of the fish senegal bichir polypterus senegalus

the authors assume this to be due to the absence of the pulp functioning as a damping layer for the stiffer dentin bruet et al used instrumented nanoindenation to measure the multilayered structure of four different organicinorganic nanocomposite material layers of the scales of the fish senegal bichir polypterus senegalus
they could show that the mechanistic origin of penetration resistance include a juxtaposition of multiple distinct reinforcing composite layers
these layers each undergo their own unique deformation mechanisms and regions of differing levels of gradation within and between material layers

a system that has to endure high amounts of stress under pressure is more effective against abrasion wear if it has a gradient in structure and properties because such a design leads to more uniform stress distribution and thus to the minimization of probability of local stress concentration which lead to the materials fatigue and failure
because a hard inflexible system will break easily under pressure and a flexible system will be easily worn off the gradient material will improve wear resistance by combining advantages of both stiffness outside and flexibility inside 42
comparison to other keratinous tissues the range of mean values of the modulus and hd obtained for the outer 4660 gpa 019041 gpa and isls 3243 gpa 013019 gpa are comparable with results obtained in other studies on the mechanical properties of keratinous materials

astbury bell state that the feather geometry and the material ultrastructure are important features in determining the mechanical behaviour of a whole feather
in the present study it is shown that despite a variation in layer thickness and overall epidermis thickness hd and stiffness measurements for all four species tested lead to similar results for the hd and modulus values for the isl and osl penetration depths of 300700 nm
perhaps this is an optimal range of values which might reflect an adaptation for sliding locomotion

surface anisotropy surface anisotropy may additionally influence redistribution and canalization of contact forces during sliding on the substrate
specific surface microstructure may lead to different abrasion quality and canalization of scratches in one particular direction
in this case generation of new scratches might even be responsible for sustaining the anisotropic microstructure for some period of time

specific surface microstructure may lead to different abrasion quality and canalization of scratches in one particular direction
in this case generation of new scratches might even be responsible for sustaining the anisotropic microstructure for some period of time
because scales of different snakes exhibit different surface microstructures both the abrasion quality and its magnitude may also be dependent on the specific geometry of the microstructure

distinct denticulations found on the ventral scales of l
californiae presumably lead to canalization of forces resulting in scratches in longitudinal direction

such longitudinal grooves generally lead to a lower initial and steady state wear if compared with transverse grooves canalization will probably not be exhibited on scales with little to no surface structure such as the ventral scales of g

such longitudinal grooves generally lead to a lower initial and steady state wear if compared with transverse grooves canalization will probably not be exhibited on scales with little to no surface structure such as the ventral scales of g
here shallow but numerous scratches without a specific orientation are observed
conclusion and outlook data of our study show for the first time the hd and eem values at different depths of the ventral skin of four snake species

conclusion and outlook data of our study show for the first time the hd and eem values at different depths of the ventral skin of four snake species
the comparison of the surface microstructure and material architecture demonstrated a gradient in material properties of the integument from a hard and inflexible outside to a soft and elastic inside
this feature is assumed to be a functional mechanism explaining abrasion resistance of the skin material an additional mechanism might be the presence of surface microstructure canalizing wear in the longitudinal direction

Article Info:
i19 Cell wall structure and formation of maturing fibres of moso bamboo (Phyllostachys pubescens) increase buckling resistance
Xiaoqing Wang1,*, Haiqing Ren1, Bo Zhang2,3, Benhua Fei2 and Ingo Burgert3,*
Published online 14 September 2011
Search Result:
the mechanical stability of the culms of monocotyledonous bamboos is highly attributed to the proper embedding of the stiff fibre caps of the vascular bundles into the soft parenchymatous matrix
owing to lack of a vascular cambium bamboos show no secondary thickening growth that impedes geometrical adaptations to mechanical loads and increases the necessity of structural optimization at the material level

here we investigate the fine structure and mechanical properties of fibres within a maturing vascular bundle of moso bamboo phyllostachys pubescens with a high spatial resolution
the fibre cell walls were found to show almost axially oriented cellulose fibrils and the stiffness and hardness of the central part of the cell wall remained basically consistent for the fibres at different regions across the fibre cap
a stiffness gradient across the fibre cap is developed by differential cell wall thickening which affects tissue density and thereby axial tissue stiffness in the different regions of the cap

to gain further insights into the mechanical performance of cell wall components and their interaction it is necessary to probe the fine mechanical details within the cell wall and correlate them with the structural features of cell wall such as microfibril angle and lignification
nanoindentation allows for a close up view on the mechanical properties of plant cell walls at the submicrometre level and has been applied in a variety of studies on wood bamboo and other cellulosic fibres by this technique not only the mechanical properties of secondary wall layers but also the mechanical variability within cell wall layers can be evaluated owing to the small indent size in the order of 100 nm in the present work we used nanoindentation technique to investigate the cell wall mechanical properties of fibres within the vascular bundles of an immature bamboo culm in which cell wall thickening and lignification have not been finalized yet

in parallel confocal raman microscopy was used for imaging the chemical composition of the vascular bundles and the cellulose and lignin distributions in different cell wall regions were visualized with high spatial resolution

raman mapping the sample blocks for raman mapping were embedded with water soluble polyethylene glycol peg 2000
specifically the sample blocks were firstly dipped in a 11 solution of peg and water at 60c for 60 h and were then kept in pure peg at 60c for 24 h prior to hardening by cooling down at room temperature
approximately 6 microm thick cross sections were cut from the embedded blocks using a rotary microtome leica rm2255 germany

mechanical tests were performed with a nanoindenter hysitron ubi1 usa which was mounted on a vibration isolation table to minimize the disturbance of environmental vibrations
a three sided pyramid diamond indenter tip berkovich type with a radius of curvature of about 150 nm was used to image and indent the cell wall
the indenter tip was loaded in a force controlled mode to a peak force of 180 micron at a rate of 36 micron s1 then held at constant load for 6 s and further unloaded at a rate of 36 microns1

a three sided pyramid diamond indenter tip berkovich type with a radius of curvature of about 150 nm was used to image and indent the cell wall
the indenter tip was loaded in a force controlled mode to a peak force of 180 micron at a rate of 36 micron s1 then held at constant load for 6 s and further unloaded at a rate of 36 microns1
the longitudinal reduced elastic modulus ie indentation modulus and the hardness were calculated from the recorded loaddisplacement curves according to the method developed by oliver pharr three positions on the fibre cap with different distances to the vessel elements were probedinner thick walled fibres adjacent to the vessels fibres in a transition region with thinner cell walls and larger lumina and relatively thin walled fibres in a region towards the periphery of the fibre cap

the indenter tip was loaded in a force controlled mode to a peak force of 180 micron at a rate of 36 micron s1 then held at constant load for 6 s and further unloaded at a rate of 36 microns1
the longitudinal reduced elastic modulus ie indentation modulus and the hardness were calculated from the recorded loaddisplacement curves according to the method developed by oliver pharr three positions on the fibre cap with different distances to the vessel elements were probedinner thick walled fibres adjacent to the vessels fibres in a transition region with thinner cell walls and larger lumina and relatively thin walled fibres in a region towards the periphery of the fibre cap
in order to create a profile of indentation modulus and hardness across the cell wall of fibres a close spaced indenting was adopted

the longitudinal reduced elastic modulus ie indentation modulus and the hardness were calculated from the recorded loaddisplacement curves according to the method developed by oliver pharr three positions on the fibre cap with different distances to the vessel elements were probedinner thick walled fibres adjacent to the vessels fibres in a transition region with thinner cell walls and larger lumina and relatively thin walled fibres in a region towards the periphery of the fibre cap
in order to create a profile of indentation modulus and hardness across the cell wall of fibres a close spaced indenting was adopted
comparative studies with different spacing widths indicated that the loaddisplacement curves are only marginally affected by the indent density in bamboo fibre cell walls data not shown

in order to create a profile of indentation modulus and hardness across the cell wall of fibres a close spaced indenting was adopted
comparative studies with different spacing widths indicated that the loaddisplacement curves are only marginally affected by the indent density in bamboo fibre cell walls data not shown

fibres adjacent to the surrounding parenchyma were not selected for nanoindentation tests as their cell walls were regarded as being too thin for reliable indents

fibres adjacent to the surrounding parenchyma were not selected for nanoindentation tests as their cell walls were regarded as being too thin for reliable indents
after indentation an imaging scan of the fibre wall was performed by means of in situ scanning probe microscopy to evaluate the position and quality of indents
owing to the embedding treatment the nanoindentation experiments could not be conducted on native wet samples

although being measured at a different position of the fibre cap these characteristic fibre images fit well to the fibres observed by raman imaging figure 4
the images also denote the pattern of indents in the cell walls allowing the creation of a profile of indentation modulus and hardness of the cell wall from the lumen to the middle lamella figure 5b
the course of the indentation modulus from the lumen to the middle lamella showed the same trend for the three positions

the images also denote the pattern of indents in the cell walls allowing the creation of a profile of indentation modulus and hardness of the cell wall from the lumen to the middle lamella figure 5b
the course of the indentation modulus from the lumen to the middle lamella showed the same trend for the three positions
close to the lumen and close to the middle lamella the indentation modulus was lower than in the middle of the fibre cell wall

the inner cell walls of fibres in all regions had a consistent hardness of approximately 400 mpa
a large variation of hardness values was observed for indents close to the lumen and the middle lamella which is probably owing to the interference from the adjacent lumen and the middle lamella during the indentation process
discussion vascular bundles are a fundamental structural component of the bamboo culm as they provide mechanical support and waternutrient transport

however the outermost layer similar to the s1 layer in wood of the secondary walls of the fibres showed a rather large microfibril angle figure 4d which was also observed by parameswaran liese where fibrils are oriented at an angle of 50 with respect to the fibre axis
the nanoindentation tests for measuring a profile of indentation modulus and hardness across the individual fibre cell walls were in good agreement with the raman analysis of cellulose orientation figure 5
the cell wall stiffness showed essentially the same trend from the lumen to the middle lamella for the fibres of the three positions with higher stiffness in the middle and lower stiffness close to the lumen and middle lamella

the lower stiffness of the inner cell wall region is likely to be an artefact because of indenting close to the lumen of the fibre
it has been shown that the indentation modulus and hardness are highly dependent on both cellulose microfibril angle and lignin level of plant cell walls however the degree of lignification seems to have a minor influence on the cell wall stiffness of bamboo fibres which can be explained by the almost longitudinal orientation of the cellulose fibrils owing to lack of a vascular cambium arborescent monocotyledons show no secondary thickening growth that impedes geometrical adaptations to mechanical loads and increases the necessity of structural optimization at the material level
one crucial aspect for the mechanical stability of the trunk is the proper embedding of the stiff fibre caps of the vascular bundles into the soft parenchymatic matrix

it has been shown that the indentation modulus and hardness are highly dependent on both cellulose microfibril angle and lignin level of plant cell walls however the degree of lignification seems to have a

minor influence on the cell wall stiffness of bamboo fibres which can be explained by the almost longitudinal orientation of the cellulose fibrils owing to lack of a vascular cambium arborescent monocotyledons show no secondary thickening growth that impedes geometrical adaptations to mechanical loads and increases the necessity of structural optimization at the material level
one crucial aspect for the mechanical stability of the trunk is the proper embedding of the stiff fibre caps of the vascular bundles into the soft parenchymatic matrix
in a previous study on a specific fibre cap type of the palm w

in a previous study on a specific fibre cap type of the palm w
robusta a gradual decrease in stiffness across the fibre capwas observed to avoid stress discontinuities between the stiffening elements and the soft parenchyma tissues which was interpreted as an adaptation to the required mechanical constraints under the given growth conditions the stiffness gradient in the fibre cap of the palm is developed at two hierarchical levels on the one hand at the tissue level owing to a decrease of tissue density cell wall thickness/cell size from the inner to the outer region and on the other hand at the level of the cell wall assembly by regulating the level of lignin and its composition
in the case of the palm fibre cap this affects the axial tensile stiffness because the cell walls possess a high microfibril angle and therefore the axial material properties of cell walls are highly dependent on the shear stiffness and strength of the matrix in terms of the investigated bamboo culm the results do not indicate changes in cell wall properties across the fibre cap

in the case of the palm fibre cap this affects the axial tensile stiffness because the cell walls possess a high microfibril angle and therefore the axial material properties of cell walls are highly dependent on the shear stiffness and strength of the matrix in terms of the investigated bamboo culm the results do not indicate changes in cell wall properties across the fibre cap
the stiffness and hardness of the central part of the cell wall remain basically consistent for the fibres at the three positions across the fibre cap
as shown by the raman imaging the fibre cell walls of moso bamboo have almost axially oriented cellulose fibrils which impedes a regulation of axial tensile stiffness by the degree of lignification

Article Info:
i20
Evolutionary optimization of material properties of a tropical seed
Peter W. Lucas1,*, John T. Gaskins2, Timothy K. Lowrey3, Mark E. Harrison4,5, Helen C. Morrogh-Bernard5, Susan M. Cheyne5,6 and Matthew R. Begley7
Published online 25 May 2011
Search Result:
here we show how the mechanical properties of a thick shelled tropical seed are adapted to permit them to germinate while preventing their predation
the seed has evolved a complex heterogeneous microstructure resulting in hardness stiffness and fracture toughness values that place the structure at the intersection of these competing selective constraints analyses of different damage mechanisms inflicted by beetles squirrels and orangutans illustrate that cellular shapes and orientations ensure damage resistance to predation forces imposed across a broad range of length scales

eden prairie mn usa equipped with a berkovich diamond tip of calibrated head radius of approximately 100 nm a similar set up to recent studies on wood specimen surfaces were incrementally polished down to 005 microm alumina prior to testing
for all tests modulus and hardness over displacement data were made using the continuous stiffness method to a depth of 1 microm at a strain rate of 01 s 1
adrift correction factor as determined by a 50 s hold step at 10 per cent of the maximum load during unloadingwas applied to the data to account for thermal drift during the test

adrift correction factor as determined by a 50 s hold step at 10 per cent of the maximum load during unloadingwas applied to the data to account for thermal drift during the test
measurements reported were the average of at least 15 indentations with 55 tests being completed on the randomly oriented fibres of zone ii to ensure that variations in local material properties did not influence the global average
the activity of beetles the larger of two coccotrypes spp

measured mechanical properties the structure of the seed shell is shown in schematic view in figure 1 with micrographs in figure 2
it possesses both thin elongate fibres and short stubby cells but for either the cell wall comprises approximately 95 per cent of the cell volume leading to an overall density that is large compared with most woods such a high density elevates the modulus and hardness
the spatial distribution of cellular shape and orientation leads to an important differentiation in mechanical properties

the shell has a nearly spatially uniform hardnessthe highest recorded for nut shells
while the elastic modulus varies more than hardness the major influence is whether loading on the fibres is axial or transverse table 1
clearly in zone ii the mean modulus is a statistical average over a range of orientations

the tunnel surfaces are smooth and plastically smeared figure 3ab
once inside the shell the females lay eggs and die their mandibles being completely worn down during a single entry severe wear is inevitable since the hardness of insect mandibles even with metal impregnation is comparable to that of the m
hatched larvae then feed on the seeds fatty endosperm develop and mate

hatched larvae then feed on the seeds fatty endosperm develop and mate
the force level required for the beetle to tunnel using small repetitive gouges figure 4 can be estimated by the force to permanently damage a brittle surface via indentation
this force is described by where h is the shell hardness e is its elastic modulus and rm is beetle mandibular radius

the force level required for the beetle to tunnel using small repetitive gouges figure 4 can be estimated by the force to permanently damage a brittle surface via indentation
this force is described by where h is the shell hardness e is its elastic modulus and rm is beetle mandibular radius
we used nanoindentation a technique involving contact radii of the order of micrometres to obtain the hardness and elastic modulus of the shell

this force is described by where h is the shell hardness e is its elastic modulus and rm is beetle mandibular radius
we used nanoindentation a technique involving contact radii of the order of micrometres to obtain the hardness and elastic modulus of the shell
with the modulus taken in the direction of indentation to the tunnel sidewall e 7 gpa one predicts f 015 n for completely worn mandibles rm 03 mm

we used nanoindentation a technique involving contact radii of the order of micrometres to obtain the hardness and elastic modulus of the shell
with the modulus taken in the direction of indentation to the tunnel sidewall e 7 gpa one predicts f 015 n for completely worn mandibles rm 03 mm

by comparison the maximum experimental mandible force for various larger beetle species has been reported in the range of f 0104 n this calculation strongly suggests that shell properties lie at the minimum required to prevent insect boring except by the most specialized species

analysis of large animal predation orangutans pongo pygmaeus large fruit eating great apes whose feeding behaviour has been studied intensively eat m
parviflora seeds at a number of study sites in borneo and sumatra in sabangau central kalimantan adult orangutans spent a mean 133182% of their total monthly feeding time consuming these seeds which are the hardest item they are known to consume observations of orangutan feeding suggest two mechanisms for large animal predation ie two methods to crack the seed by biting
in the first method observed in wild orangutans in sabangau the seed is placed between jaws with the germination band running vertically from bottom tooth to top tooth the crack probably originates from the plug which serves as a pre crack aligned with the bite force and pointing into the germination band figure 4

at the same time when combined with the intricate shell design the fracture toughness is large enough to deter large predators
predation by small consumers is discouraged by high cellular density that produces a relatively uniform obstacle to penetration at small scales
the seed has thus evolved highly adapted shell architecture with optimal properties in the sense that they are tailored to balance competing performance constraints to ensure survival there appears to be little room for property deviation

the calculation of internal pressure required to crack the shell suggests that shell thickness and stiffness are maximal thicker stiffer shells would require pressures that significantly exceed known turgor pressures preventing germination
while softer more compliant shells would facilitate germination by requiring lower turgor pressures to crack the shell open even small reductions in hardness and stiffness would leave the shell prone to burrowing by beetles
similarly the fracture toughness of the shell is low enough to permit germination but not sufficiently low as to allow chipping by rodents

to resist the attack of large animals the shell needs to be thick stiff and most of all tough rather than maximizing these properties outright the shell has evolved a complex structure that raises these properties to upper limit indicated by the need to crack open during germination
the calculations outlined above emphasize the critical role of the germination band it is equally as hard as the surrounding shell which is the principal property hindering burrowing by very small predators
at the same time this band has a lower fracture toughness to allow for germination while being narrow enough to prevent chipping by medium sized predators

since this is considerably less than the size of the plug linear elastic fracture mechanics is justified
chipping during gnawing the analysis of chipping in the main text is based on previous work that analysed the case of a sharp tool indenting a surface with a nearby free edge wherein a chip is formed between the point of contact and a distance h from the free surface the scaling in the chipping equation main text has been validated for a broad range of geometries with the dimensionless factor b accounting for the angle of the loading axis to the surface the interior angle of the cutting tool indenter and the angle of the chip corner
this factor lies in the range of b 515 with the smallest values corresponding to sharp tools and shallow inclinations of the indenter relative to the surface and the largest values for blunter tools and nearly normal indentation of the surface

chipping during gnawing the analysis of chipping in the main text is based on previous work that analysed the case of a sharp tool indenting a surface with a nearby free edge wherein a chip is formed between the

point of contact and a distance h from the free surface the scaling in the chipping equation main text has been validated for a broad range of geometries with the dimensionless factor b accounting for the angle of the loading axis to the surface the interior angle of the cutting tool indenter and the angle of the chip corner

this factor lies in the range of b 515 with the smallest values corresponding to sharp tools and shallow inclinations of the indenter relative to the surface and the largest values for blunter tools and nearly normal indentation of the surface

this range is used to compute the range of chip sizes cited in the main text