Organic Software Framework for an Adaptive Integrated Knowledge Repository

David J. Lubliner
New Jersey Inst. of Technology
Lubliner@njit.edu

George R. Widmeyer
New Jersey Inst. of Technology
Widmeyer@njit.edu

Fadi P. Deek
New Jersey Inst. of Technology
Fadi.Deek@njit.edu

Abstract

The goal of the research was to design and build an organic knowledge repository that had the inherent ability to grow and evolve over time. A design artifact – Constructivist Unifying Baccalaureate Epistemology (CUBE) – was designed and tested to validate the efficacy of this approach. Two models that facilitate this organic nature of the system were developed for this research; a Knowledge Weighting Model and the Aggregation-Integration-Master model. A Semantic Web ontology integrated with the W3C Resource Description Framework was used to create a concept space that offers a unified view of the discipline. The two goals of this approach were validated. First, students and faculty were able to enter information that dynamically altered the organization and structure of the knowledge repository. Second, students utilizing the Integrated Knowledge Repository developed a more complex understanding of the interconnected nature of the materials linking a discipline.

1. Introduction

The dual goals of designing an organic system that evolves over time and creating a knowledge repository that facilitates knowledge exploration was the focus of this research. One of the challenges educators have in the computer science and engineering fields is to ensure students graduate with an understanding of the interconnected nature of concepts in a discipline that allows them to solve unstructured problems. Technology evolves and so must the learning environment. Incorporating new materials that track the evolutionary growth of knowledge in a discipline is a daunting if not an almost impossible task for any one instructor. This work is a preliminary step in creating a framework and learning system to provide the means for faculty, students and domain experts to collaborate in creating a dynamic knowledge repository. It has some of the features of a wiki, a collaborative environment where a group of individuals iteratively update the knowledge base, but with an added quantitative framework. This framework is constructed using two new models created for this purpose; the Knowledge Weighting Model (KWM) and Aggregation-Integration-Master (AIM) model that facilitates the evolutionary growth of the system. A voting-ranking system, incorporated into the KWM, weighted by the expertise levels of the participants; faculty and domain experts carry higher weights than students, thus ensuring the high quality of the content. The solution was to develop a design artifact Constructivist Unifying Baccalaureate Epistemology (CUBE) that integrated concepts spanning a field of study that facilitated the evolutionary growth of the system as new information was integrated into the field of study.

A constructivist learning framework was selected to facilitate the creation of a dynamic learning environment where learners actively construct a representation of concepts, integrating information from multiple sources [4, 8, 9]. It has been hypothesized [14] that “Ultimately the development of content knowledge bases that integrate content across multiple courses within a degree program is an expected evolution.” The task of integrating these distinct pieces of the puzzle is usually the responsibility of the learner. The intent of this research is to provide a structure, several models, and a prototype knowledge repository to realize this goal of creating a dynamic integrated learning environment spanning an entire discipline.

In order to validate this approach a design artifact, called CUBE, was developed to integrate the materials from multiple college courses. CUBE is a dynamic environment that incorporates student input to ensure the evolution of the
knowledge base. This includes a semantic web framework [2] and Conceptual Clustering to create a concept space with conceptual threads linking concepts. A generic structure has been developed to allow other disciplines to utilize this framework. The central hypothesis is that students utilizing the Integrated Knowledge Repository (IKR) will develop a more complex understanding of the interconnected nature of the materials linking a discipline than those who take conventional single topic courses.

In order to realize this interconnected knowledge repository, the current constructivist learning environment was extended to incorporate a formal relationship between the building blocks of knowledge formation. The evolutionary components of factual knowledge, conceptual knowledge, procedural knowledge and meta-cognitive knowledge presented in Bloom’s revised taxonomy [3] are integrated into a Knowledge Weighting Model (KWM). The factual elements, course materials, form the basic elements that are interconnected with a concept-weighting structure that forms an integrated conceptual knowledge. Procedural knowledge is constructed using a knowledge map that displays the skills and algorithms as they evolve from basic to more sophisticated applications in a discipline. Finally, the meta-cognitive component captures the structure of a subject matter as cognitive tasks that are correlated using a field relevance structure that ties together the conceptual underpinning of a discipline.

An organic framework forms the foundation of the knowledge repository. Individuals, Students-Faculty-Domain experts (SFDe), enter semantic terms that represent current concepts. The SFDe's provide links to content that tie together the field. Then they vote on the importance relevance of those semantic terms and links. Conceptual Clustering is used to rank the interdependencies and display them as an evolving concept map. The CUBE design artifact incorporates all these components.

Section 6 of the paper presents both qualitative and quantitative evaluation of the artifacts. This research presents a clear and verifiable solution to the problem of engaging students in “deep learning” of course materials and followed a rigorous research approach with both formative and summative evaluation. While this paper focuses on the technological details of the design, we have presented the results in other papers for an education audience [10].

The innovative features of this research are twofold. The first is the organic-adaptive nature of the design. New content; semantic terms, relevant links to new materials, consensus voting and conceptual threads linking concepts spanning a discipline are dynamically displayed utilizing the knowledge weighting model (KWM). The KWM uses conceptual clustering that determines the topics and links that all voting members agree upon, by frequency and correlation strength, as structurally important to the field of study. The second new innovative feature is the research, described in Section 6, that presenting information to students as a continuum of thought, with tools that allow them to independently explore these threads, leads to a quantitative improvement in overall comprehension of the cohesive concepts that link a discipline. Phase II of the project, which will begin shortly, uses these tools to link multiple universities and will investigate whether these collective tools can benefit from a larger community’s feedback.

This research is important to the field of Information Systems since there is currently no system or model that provides the organic structure to facilitate the co-evolution of a knowledge repository where the system is designed to support changes in the user and the knowledge structures. In addition, the CUBE design artifact can be used as a model to integrate information in other fields of study.

2. Theoretical Framework

Three approaches to learning that have evolved during the last century; Learning as response strengthening, learning as knowledge acquisition and learning as knowledge construction. [9] The first approach has the learner passively receiving reward and punishments, such as drill and practice, simple response and feedback. The second has students placing new information in long term memory. The learner still passively acquires information from the teacher who presents information in textbooks and lectures. Knowledge is a commodity transmitted from the teacher to the learner. The third approach, learning as knowledge construction, is based on the concept that learners actively construct a knowledge representation in working memory.

Piaget’s theory of cognitive development in children [13] postulated a sequence of four qualitatively distinct stages of intellectual development; Sensor-motor, Preoperational, Concrete operations and Formal operations He
believed that “the learner must be active; he is not a vessel to be filled with facts…Learning involves the participation of the learner”. Creating an environment designed to allow students to explore and independently navigate tendrils of interconnecting concepts will empower and enhance their construction of a more cohesive understanding of interconnected facets of a discipline. Later in the 1900’s, Vygotsky’s [15] Zone of Proximal Development (ZPD) stated that the potential for cognitive development depends on social development. Skills that can be developed in collaboration with their peers exceed those which can be attained alone. Later in the 1990’s, theories of human learning in realistic settings [9] emerged, based on the belief that the learner is the sense-maker and the teacher is the cognitive guide who provides guidance and modeling on authentic academic tasks. The CLE provides a framework for designing and building the third approach.

2.1 Constructivist Learning Concepts

Constructivist learning theory was used as a framework for the design of the organic knowledge repository. Constructivism postulates that learners construct knowledge for themselves. Individually and socially they construct meaning as they learn. The goal of this research was to build on the core tenets of constructivism by developing two models, integrated into a knowledge repository that facilitates a student’s exploration and knowledge synthesis, to more effectively integrate knowledge spanning a discipline than current instructional models.

Figure 1 shows the conceptual approach for the design of the knowledge repository. At each step in the process the constructivist tenet is followed by the model or technique used to create the design artifact, CUBE.

The first step was to create a model to instantiate the Constructivist tenet that is essential for knowledge integration, which is that students reflect on their learning activities and observations. The Aggregation-Integration-Master (AIM) model builds upon Mayer’s Selection-Organization-Integration (SOI) knowledge construction model [11]. The Aggregation-Integration-Master model, developed as part of this research, extends the SOI to a larger domain.

Figure 1 Flowchart of Constructivist Learning Framework with Embedded Organic Components
It is postulated that effective knowledge integration and comprehension is only truly effective if it correlates all components of a discipline into a cohesive whole. This also incorporates the Active tenet that students develop their own interpretation that aids in knowledge synthesis, by allowing them to iteratively investigate information pathways. The next constructivist step is that learners articulate what they have learned. This is accomplished by integrating a Semantic Web ontology [2, 5] “set of formats and languages that are used to find and analyze data on the web,” integrated with the W3C Resource Description Framework to create a concept space that offers a unified view of the discipline. A ranking/voting mechanism facilitates the constructivist intent that users articulate what they have learned and collectively provide input to the repository.

The fourth constructivist tenet - Authentic - is supported by the use of case studies. This insures that students integrate real world examples to ground their overall information synthesis by exploring the complexities of practical problems that don’t always have finite linear solutions. The real test of an individual understanding, in a particular field, is the ability to incorporate core algorithms and knowledge to solve new unstructured problems.

Finally, the students are encouraged to cooperate in their problem solving. They suggest resources and vote on the characteristics and value of those resources to their own understanding. The voting is anonymous and the knowledge weighting model, explained in Section 4.1, provides a consensus evaluation of the suggested resources. This provides feedback to individual students and motivates further participation.

3. Organic Design Principles

Organic Design postulates that software can evolve to meet the needs of the environment. Table 1 defines four new design principles that should be incorporated into any organic design to facilitate its evolutionary growth. The static nature of most systems limits their functionality to the original design parameters, quickly becoming obsolete and requiring a never-ending array of revisions and service packs. Biological systems evolve through mutation, not all beneficial, responding to input from their environment to adapt or eventually die.

An underlying concept in biology, in the publication, Science [6], describes a model that affects alleles in influencing adaptation to complex environments. All organisms strive to adapt to changing environments. The Science article describes biological adaptation to rapidly changing environments. Software designers face similar challenges in a rapidly changing landscape. These design principles, described in Table 1, mimic evolutionary adaptive design.

<table>
<thead>
<tr>
<th>Item</th>
<th>Name</th>
<th>Design Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Symbiotic</td>
<td>Feedback and Voting Mechanisms to enhance the symbiotic relationship</td>
</tr>
<tr>
<td>2</td>
<td>Emergent</td>
<td>Co-evolution of software and user knowledge</td>
</tr>
<tr>
<td>3</td>
<td>Ecological</td>
<td>An adaptive user interface that reflects a changing environment</td>
</tr>
<tr>
<td>4</td>
<td>ADaptive</td>
<td>An adaptable architecture</td>
</tr>
</tbody>
</table>

Table 1 Proposed Organic Design Principles

The symbiotic relationship (Item 1) is essential in this evolutionary phase (Item 2) of organic system design. In biological systems organic evolution is in response to changing ecological stimuli, environmental and competition among species. In computer evolution the system has to respond to environmental changes, network / software evolution and the evolving needs of the biological systems accessing and interfacing with the computer architecture. One recent project, the Tortola Project [7] has proposed a layer between hardware and software “a virtual interface that enables hardware and software to communicate and to solve problems together.”

The project focuses on adapting to power, reliability and security. An organic system design must adapt to the needs of the biological systems and develop a symbiotic relationship, not just adapting to evolving hardware requirements.

3.1 Generic Organic Structure

As illustrated in fig. 2 the initial configuration (4-sided figure) evolves where the successive linear incremental changes, second instantiation, (8-sided figure) maintains some of the original characteristics. As the evolution continues, non-
linear, core components may still be recognizable but can evolve and generate non standard configurations (11-sided and not completely closed) referring to possible new content from external disciplines that might be incorporated.

3.2 Organic CUBE Components

The organic design principles are interwoven into the architecture of the CUBE artifact. Students-Faculty-Domain experts (SFDe) provide semantic terms and links, generating W3C RDF triplets and dynamic relationships (Emergent principle). SFDe’s then vote on the importance/weighting of individual concepts and relationships (Ecological & Symbiotic principles). Finally a conceptual map, fig. 3, reflects changes in the linkages between concepts (aDaptive principle). Those elements and relationships with the highest probabilities are incorporated into the knowledge map as new data / knowledge is entered into the system.

3.3 Learning Objects

A learning object is “a digital self-contained and reusable entity, with a clear educational purpose, with at least three internal and editable components: content, learning activities and elements of context.”[17]. A 3D concept map links learning objects and content generated by the Student-Faculty-Domain Experts’ fig. 4.

4. Design Models

Two design models were developed as part of this research – the knowledge weighting model and the aggregation-integration-master model.

4.1 Knowledge Weighting Models (KWM)

In order to integrate knowledge that spans an entire discipline, there has to be a well-defined model to weight the individual course topics and concepts, the common elements that exist between the materials, the correlation weights for the interdependence of the variables and the evolving relevance of existing and new material to the overall growth of the discipline. One measure of knowledge...
weighting is field relevance between the materials, the correlation weights for the interdependence of the variables and finally the evolving relevance of existing and new material to the overall growth of the discipline. Another is a quantitative measure of correlation between concepts that is provided by input from faculty and students by voting on correlation weights.

In order to validate the Knowledge Weighting Model, a machine learning paradigm for classification using conceptual clustering was utilized. Conceptual clustering uses the inherent structure of the data that drives cluster formation and a description language. It determines classes with common characteristics extracted from large amounts of data.

There were 10-20 students per course, generating 3-7 links per topic over a 15-week semester. The classes generated approximately 1100 links per course. In addition, students ranked their top links. Five choices per topic over 15 weeks times 10-20 students yields an additional 1000 ranking data points. Therefore, there were (1100 + 1000) or, approximately 2100 data points collected per course that are used to generate the concept map.

4.2 Aggregation-Integration-Master (AIM) Knowledge Construction Model

The Aggregation-Integration-Master (AIM) model builds upon the Selection-Organization-Integrate (SOI) knowledge construction model of Mayer [11]. The Aggregation–Integration-Master model, developed as part of this research, extends the SOI to a larger domain. It is postulated that effective knowledge integration and comprehension is only truly effective if it correlates all components of a discipline into a cohesive whole.

This provides an iterative process where relationships and links between concepts provide mechanisms to navigate interrelated concepts. The model satisfies the tenets of the constructivist approach to knowledge formation where students are the knowledge makers and more effective integration and visualization of meta-cognitive data linking is continually evolving. In addition, the knowledge repository itself continues to evolve, integrating new links that ensure the information is timely and relevant.

In order to validate the effectiveness of this approach, KWM and AIM models, a prototype was designed, built and tested utilizing students spanning several semesters in multiple courses.

5. CUBE Design Artifact

The Integrated Knowledge Repository (KR) aggregates course materials of N courses with associated concept maps. The CUBE, (Constructivist Unifying Baccalaureate Epistemology), KR incorporates constructivist features is a schema for enhancing learning and knowledge formation.
5.1 Application Layer

The application layer provided a macro view of the knowledge repository. The research started with four courses. Courses were represented as planes on a three-dimensional cube, fig. 7. Through testing, focus groups and semi-structured interviews, it was determined that providing a simple mechanism/user interface for students to visualize the relationship between courses and the sequence in the curriculum facilitated the overall view of the concept space. The simple user interface allowed students to select individual lectures. A concept map provided a detailed navigation path between related concepts; this is the key to the application layer, fig. 3. There are linkages between the concepts in a course and between concepts across courses. The weights on the linkages are completely defined by student input.

![Figure 7 Generic Visualization structure](image)

5.2 Knowledge Layer

The cube pointers provide successor/predecessor links that provide the ability to traverse courses/concepts that span a discipline. When defining a relationship, a string is defined with a syntax that provides for easy parsing of the message. Square brackets indicate the scope of N primary successor/predecessor relationships. [CA 125: Suc 126: Pred 124] . Curved brackets refer to core materials that are outside the particular field of study (Core CHEM 125: Physics 120). Links un-encapsulated refer to added links with relationships that have not clearly been defined or voted on.

The CUBE header structure is modeled after the Transmission Control Protocol (TCP) message headers to allow for future expansion of the message headers that define the relationships between entities. It is designed to allow for the addition of multiple predecessor/successor relationships. An additional feature is the ability to tie information from related disciplines to create a more global knowledge structure, linking core concepts that provide the foundation for multiple disciplines; i.e. math, physics, chemistry etc. fig. 8.

![Figure 8 Cube linkages between concepts](image)

5.3 CUBE Data

The initial configuration of the Cube knowledge repository was populated using content traditionally used in single content courses; i.e. course outlines, objectives, lectures, examples and assignments that assessed their skills. Students, faculty and domain experts (SFDe), during the two year testing, provided semantic terms derived from individual course topics. In addition SFDe's provided links to new SFDe's then voted on the importance and relevance of these linkages that populated the conceptual maps, fig. 10.

![Figure 9 Content links and relationships](image)
6. Evaluation

The CUBE artifact was evaluated using a case analysis rubric. Case analysis problems are often complex, interdisciplinary problems. These problems engage the learners in understanding and resolving issues, rather than remembering them. It requires learners to critically analyze situations, identify issues and assumptions and engage in reflective thinking. The levels of learning and thinking engaged by this process are at a much deeper level [8].

The system provided students with a series of cases in the field of Computer Technology. The students’ responses were analyzed using a rubric to determine if the Integrated Knowledge Repository approach was effective in facilitating a more complex understanding of the interrelated nature of the discipline spanning multiple courses. The assessment criteria categories include: Quality of Information Sources Cited (QISC), Constraint Analysis (CA), Feasibility (F) and Relevance of Implications (RI), as summarized in Table 2.

There were two hypotheses that were tested to evaluate the CUBE knowledge repository. H1: “Students using the Integrated Knowledge Repository (IKR) will have a more positive perception of the learning process than those who use conventional single course teaching paradigms”. H2: “Students utilizing the IKR will develop a more complex understanding of the interconnected nature of the materials linking a discipline than those who take conventional single topic courses.”

6.1 Qualitative Questionnaire: Evaluate Users’ Perceptions of the CUBE Design

This questionnaire evaluated students’ experiences using the CUBE knowledge repository. The questions were broken down into two categories. The first was the students’ perceptions of the system and user interface. The second group addresses content integration; i.e. the efficacy of utilizing this approach as it pertains to knowledge acquisition and cohesion of concepts spanning a discipline.

The results of the qualitative questionnaire indicated that students using the Integrated Knowledge Repository, CUBE, had a more positive perception of the learning process than those who use conventional single course teaching paradigms. It indicated that students believe, by over 90%, that the CUBE system will enhance their comprehension of subject matter.

6.2 Quantitative Analysis

A quantitative approach was used wherein students were given an exam that covered material spanning multiple courses. The goal was to determine if students attained higher scores using the knowledge repository instantiated by the CUBE artifact. Three groups of students were given the same questionnaire, in the same order, using the same written instructions to reduce tester bias. The intent was to ensure an accurate knowledge baseline before administering the exam to the treatment group.

- **Baseline data**: students majoring in the Electrical and Computer Engineering, ECET, were given the questionnaire as a baseline to determine the skill level of students in courses ranging from their sophomore to senior years. It was determined that students in this related discipline would have similar skills. (37 Electrical and Computer Engineering, ECET, students)
- **Control group**: A control group, Computer Technology students, consisted of subjects who had equivalent or similar characteristics as the experimental group at the start of the study. The latter group received the treatment or independent variable being investigated while the control group received a placebo or another treatment. In this situation the control group were students who did not use the CUBE system. (19 Computer Technology, CPT, Students)
- **Treatment Group**: Students using the CUBE system were evaluated to test whether the hypotheses could be substantiated. (34 Computer Technology, CPT, Students)
- **Total N (37+19+34) = 90**
Table 3 Summary of research population and test results

- The similar means of the baseline group 50.0 and the control group 53.68 indicated that there was a similar level of common knowledge that could serve as a baseline for the knowledge repository.
- The means of the treatment test scores were (79.41-53.68) or 25.73 points higher. This indicates a clear improvement in scores, utilizing the knowledge repository.
- These courses were taught by four separate faculty members to reduce researcher bias.
- Using the two data sets, control and treatment groups, each characterized by its mean, a t test was used to determine whether the means are distinct, with underlying distributions that are normal. A t-test value of 0.0298 indicated that the null hypothesis could be rejected; there is a significant difference.
- CPT 493 and CPT 493H, Medical Informatics, were both taught using the knowledge repository, with the same instructor spanning two semesters. The 493H class was a hybrid course, 50% face-to-face instruction and 50% online instruction. The intent was to evaluate the possible effects of an online environment on the IKR. The results showed a 4.75 point increase for those students using the hybrid course. This suggests that for students who are already comfortable using a web based learning environment, the IKR may further amplify the positive learning benefits. The feedback from the students’ impressions of the system was very positive. The ability to explore additional material, which helps clarify the concepts covered, appeared to empower them as active participants in the learning process. In addition, for instructors who may have limited time to explore and add new course content, the quality of the course would be richer with greater depth given the additional content provided by the students. Regarding assisting students with homework, students added links that provided graphical tools which allowed them to explore and understand the problem solving process in greater depth than would normally occur.

7. CONCLUSION

The goal of this research was to design and test a Knowledge Repository for a discipline that had the inherent ability to organically evolve. Students, Faculty and Domain Experts dynamically added to the content of the CUBE knowledge repository over the course of two years. The authors believed that incorporating constructivist learning theories, which are based on the assumption that students need to be partners in the learning process with some level of control to navigate and integrate information tailored to their individual learning styles, should enhance the students’ overall comprehension of the complexities that integrate a discipline. Both of those objectives appeared to be met. The knowledge repository successfully integrated new materials. Students ranked and voted on concepts. Their input on questionnaires indicated their positive perception of the system and the approach to facilitate exploration of concepts. Students utilizing the CUBE knowledge repository showed a substantial increase in test scores on a standardized exam over students taking the conventional single course method. Utilizing a web-based tool that allows students to be active participants in constructing meaning by integrating course materials spanning a discipline created a more engaging and effective learning environment. The research reported in this paper clearly seems to indicate the potential of the CUBE approach to an integrated knowledge repository.

What wasn’t apparent before this study was that there could be a quantifiable increase in understanding of a discipline by students if they had access to more advanced concepts and topics from subsequent courses, which they have not yet taken, at the earliest level of instruction. The ability to visualize and explore the entire discipline, even if they did not comprehend the details of the more complex concepts, provided an excitement for the discipline and an understanding of the direction of their learning experience. By knowing the purpose and direction of their current studies, not at the end of their studies in a terminal course, but
reinforced all along the way, students didn’t need the “aha moment”; they could visualize the roadmap at the outset and could traverse familiar well defined pathways reinforcing the cohesion of ideas and ensuring an integrated view of the discipline.

The limitations of this research, which was conducted during a two year period, covering four courses in Computer Technology, were that the research focused on one discipline and included only a subset of all the course content of that discipline. An optimum study would compare and contrast outcomes in several disciplines such as Liberal Arts, Basic Sciences, and Engineering etc. In addition, there is a need to explore the scalability issue where every significant course in a field of study, including basic core courses such as math, physics, chemistry etc., are integrated into the knowledge base.

This research has shown promising indications that creating an organic knowledge repository with the ability to evolve can provide individuals with a better understanding of the cohesion of concepts that interconnect a field of study. By integrating this knowledge repository across multiple disciplines in a university, students could develop a better understanding of the linkages between all of the complimentary fields of study. In addition, by expanding the scope of these knowledge repositories to several universities and then to the discipline as a whole, the scope of the integrated knowledge repository would truly represent the depth and complexity of the entire field. Eventually a global interconnected knowledge repository would encompass all fields and accelerate all knowledge formation. There is currently a compartmentalized view of information. Knowledge in one field doesn’t always quickly migrate to others. The hope is that by creating organic knowledge repositories, not only will educational paradigms evolve but transparencies between disciplines will diminish and knowledge repositories will naturally evolve to incorporate a wider breadth of information.

REFERENCES