Understanding Science Through Knowledge Organizers: An Introduction

Meena Kharatmal and Nagarjuna G.
(meena, nagarjun)@hbcse.tifr.res.in

Homi Bhabha Centre for Science Education (TIFR)

We propose, in this paper, a teaching program based on a grammar of scientific language borrowed mostly from the area of knowledge representation in computer science and logic. The paper introduces an operationizable framework for understanding knowledge using knowledge representation (KR) methodology. We start with organizing concepts based on their cognitive function, followed by assigning valid and authentic semantic relations to the concepts. We propose that in science education, students can understand better if they organize their knowledge using the KR principles. The process, we claim, can help them to align their conceptual framework with that of experts' conceptual framework which we assume is the goal of science education.

Introduction

At the turn of the last century a group of philosophers of science, popularly known as *logical positivists*, began to build the grammar of science. In the current intellectual atmosphere logical positivism is more or less considered a sin. One main reason for that being the epistemological ground of positivism, that scientific theories are grounded and logically connected in observational language, was more or less convincingly demonstrated to be incorrect. In the process the baby was thrown with the bath water. What we are trying to do here is to propose a way to fetch what was lost in that bath water, the grammar of scientific language.

It is justified to believe that science, unlike folk-lore or common sense, is more *rigidly* organized body of knowledge. Exact sciences like mathematics and physics tend to be very economical in the number of concepts used to describe the phenomena and the connectors, to express the possible relations, between the concepts. It is the objective of science to eliminate ambiguity and use concepts as precisely as possible. Scientists do this by creating an artificial language of their own, a constrained ‘natural’ language. This constrained natural language is a very small sub-set of any natural language in which science is communicated.

Since communication of science is often done in a natural language (most commonly in English) we fail to explicitly realize that the *grammar* of this scientific language is distinct from those of natural languages. Consequently when we teach science we fail to communicate its grammar. We think that the practice of science teaching also suffers due to a lack of explicit teaching program to introduce the grammar of science. With this assumption in mind we began exploring to create a teaching program based on a grammar of scientific language borrowed mostly from the area of knowledge representation in computer science and logic. What we present here are some preliminary results.

In one preliminary study we found that students encounter about 4000 concepts of biology (excluding the names of all the species of plants and animals) up to higher secondary level of education (equivalent to K12) in a typical Indian school[20]. Complexity of biological science is well known, and describing such a phenomena obviously requires a richer language. Added to this is the fact that most of biological terms are derivatives of Latin or Greek. When confronted with such a large and ‘remote’ vocabulary, biology teachers often explain the etymology and explain the formation rules of such terms explaining in terms of suffix and prefix derivatives. Another very interesting recourse that biology teachers take is the abundant use of diagrams. This does help to a large extent. Carefully illustrated diagrams communicate the precision required in science, sometimes more successfully than written words. Apart from these normally followed methods, we think, it is important to add to it an explicit teaching of the grammar of scientific knowledge. This approach, in addition to enhancing precision in science communication, will also help in improving conceptual understanding of the subject.

A grammar of a scientific language consists of a finite (not necessarily known) set of possible relations between the concepts. For example, in the statements “Ribosomes are part of a cell.”, “A cell is a structure.”, and “Rabbit is a mammal.”, we employed the relation types, ‘part of’, and ‘is a’, between the terms. Our hypothesis is that though the terms are numerous the *type of relations* are not. A set of such type of relations constitute knowledge organizers (see section on methodology) of a body of knowledge, which we assume are not only constant but few in number. Our hypothesis is: during the course of science education if students are trained to understand the scientific knowledge using the knowledge organizers, then *meaningful learning* vis-a-vis *rote learning*, as explicated by Ausubel[1], takes place. When we explored for a set of required knowledge organizers for science (or for the domain of biology) from the literature, we could not obtain with the exceptions of CYC[12], UMLS[16], any such set readily available. This indicates that there is a need to develop an authentic set of knowledge organizers for use in science education. Our research objective is to fill this gap.

The epistemological presuppositions (the working hypotheses) of this undertaking are: (1) a cognitive agent understands a new concept when relations are established between the preexisting concepts with the new concept[1, 13, 15]; (2) to educate a person therefore is to facilitate the process of establishing the relevant relations between concepts so as to align with that of an expert; (3) learning therefore involves restructuring of conceptual schemes; and (4) misunderstandings are due to mismatching between conceptual schemes between the agents. According to this approach no concept gains any meaning independent of its relations with other concepts. Thus, meaning of a concept is the network it forms with others. A central difference between the approach followed by Novak and ours is the emphasis on a minimal set of unambiguous knowledge organizers instead of using many often ambiguous relations names.

The sense of understanding used here is stronger since we are seeking that the relations between the concepts be made explicit. For example, when we look at a tree, and recognize that it is indeed a tree, is also understanding of a sort, but it is implicit. Also the term 'education' is used here in a strong sense. This does not cover the various forms of behavioral mastery, such as skills, that children learn and execute without any explicit understanding. One of the challenges in education, particularly of exact sciences, is to gradually train learners towards more and more explicit forms of representation. Formal sciences like theoretical physics, mathematics and logic, for example, are domains of discourse where procedural knowledge is declaratively stated and declarative knowledge is procedurally stated reaching a highest degree of explicit knowledge representation[9, 6].

This knowledge organization (KO) approach helps in building a framework for curriculum development, and also to understand the transformation (conceptual change) of novice into an expert. Curriculum designed using KO approach follows a principled disambiguous approach, which is used by the experts in their respective domains. Incorporating the principled/logical approach is very essential to transform a novice into an expert, which we assume, is the goal of education. Based on a comparison of a novice with an expert[2], studies have shown that an experts' knowledge structure is coherent, economical and tightly integrated, while a novice's knowledge structure is often inconsistent, ambiguous, and loosely organized. While attempting to organize knowledge, an expert starts with the core concepts, however a novice starts to organize the knowledge from periphery. The approach followed by an expert is principled i.e. logical, which is not the case with a novice.

Based on these assumptions we began to employ the basic concepts of knowledge representation (KR) and its possible use in the current endeavor. It was noted by Fisher[5] that KR helps students in order to learn better. The following is a summary of the arguments given by Fisher for using KR approach in science education. The act of creating an organized structure of ideas on paper or on a computer helps in creating a knowledge structure in the mind. KR helps in making the implicit (often fuzzy) knowledge into an explicit and precise knowledge. It incorporates cognitive and metacognitive skills, thus occurs meaning-making. KR helps students to make finer discriminations between ideas and helps to organize better. The more one practices the better one becomes at organizing and relating concepts. Structural (organized, semantic) knowledge is essential to assimilate, recall and comprehend. Structural knowledge is essential to problem solving. A collaborative task occurs on the discussions about the meanings of concepts and the relations between the students. It has also been noted that there exists significant differences between the structural knowledge of novices and experts, and hence for novices a natural part of learning is to work on their structural knowledge to make it more expert-like.

Many educational researchers have found it useful to adopt a network representation format for explicitly representing knowledge structure. There exists various methods to represent knowledge such as—concept map[15], knowledge Vee[15], Concept Circle Diagrams[5], SemNet[5], Conceptual Graphs[19]. After analyzing the concept mapping methodology, we identified several problems on the basis of the assumptions stated above, and particularly due to the use of knowledge organizers. In the traditional concept mapping methodology the relation types (linking words) such as—is-a, for e.g. are ambiguously used; too many linking words are used to express the same meaning; the hierarchies are not ordered and not validated. Hence, the graphical representation is misleading to evaluate concept maps. A criticism of concept mapping methodology is discussed separately in a working paper, *Towards Principled Approach of Concept Mapping*[10]. We find the conceptual graphs approach by Sowa[18] highly instructive and we plan to make use of this technique for representing scientific knowledge. Sowa's approach is sufficiently formal to represent knowledge in precise terms, and is comprehensive enough to use in several domains of knowledge. Based on the current wisdom in KR, we developed a modeling tool to undertake the task. The tool is called GNOWSYS (Gnowledge Networking and Organizing SYStem)[8]. After introducing the model of GNOWSYS, we introduce the model followed for representing a small domain of biology. The purpose of this communication is limited to share the approach and assumptions followed.

GNOWSYS (Gnowledge Networking and Organizing SYStem)

While designing the architecture of GNOWSYS, we kept in mind the need for drawing concept graphs, semantic nets and concept maps. Recently, several researchers used concept maps[15, 14] and SemNet[4, 5] to enhance conceptual learning in the context of science education. Most of these tools, suggested in the above citations, are essentially drawing tools, and the maps drawn by the students or experts could not be stored in an accessible knowledge base. Graphs were stored as separate files, which makes reusing a component of a graph difficult. Since the graphs were encoded in a format that is internal to the applications, it is difficult to compare two concept graphs, made by different applications and remain

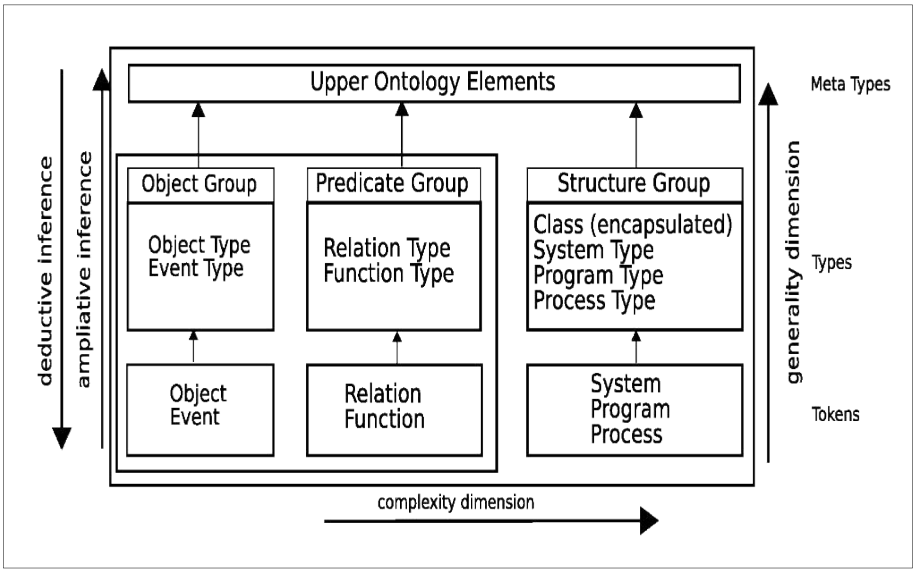

Figure 1: A diagram representing the architecture of GNOWSYS.

unsharable. The objective of matching and mismatching of concept graphs of two or more agents could not be achieved without a sharable encoding. While designing GNOWSYS these problems were kept in mind, so the graphs generated by various applications could be shared and published by the system through XML based representations schemes.

The architecture of GNOWSYS[7] is structured to accommodate different dimensions of KR such as—*generality, semantics, complexity, inference* as shown in figure 1.

Generality dimension: We come across a wide variety of concepts in our discourse either in the form of particulars or generals, and even those concepts which belong to higher levels of abstraction. These concepts need to be organized based on their order of generality. In GNOWSYS, three different levels of generality, such as tokens (for particulars), types (for generals), and metatypes (for types of types), are possible.

Semantic dimension: In GNOWSYS we can store propositions in three layers with increasing order of semantics and consistency. The first layer consists of simple propositions in the form of *well formed formulae* (WFF). In the first layer, all kinds of propositions are allowed to store without any semantic constraints. In the second layer, these WFF can then be combined with the semantic constraints, logical connectives, modalities, propositional attitudes, quantifiers etc. The system does not check for contradictions, and consistencies in this layer too, but consistency is implicit and therefore referred to as *implicit structured system* (ISS). In the third layer, the validity constraints are imposed *explicitly* and therefore gives rise to *explicit consistent systems* (ECS), which is quite similar to the experts' knowledge system. This way GNOWSYS proposes to represent knowledge of novices and experts, with the assumption that the ECS matches that of an expert and the loosely structured ISS representation with that of novice. The semantic dimension is not represented in the figure. Since it is possible to build different ontologies from a given set of vocabulary and WFF, it is possible to store multiple ontologies and epistemologies within a single or a distributed knowledge base.

Complexity dimension: Along the complexity dimension, the system supports vocabularies like simple terms, and predicates, and very complex forms like rules, arguments, axiomatic systems and other complex compositions. The basic components provided are ObjectType (OT), Object (O), RelationType (RT), Relation (R), MetaType (MT), EventType (ET), Event (E), FlowType (FT), Flow (F), which helps to store the terms, propositions and procedures. Complex compositions are provided by the structure groups consisting of ProcessType (PT), Process (P), StructureType (ST), Structure (S), Encapsulated Class, Programs and ProgramType.

Inference dimension: Epistemic values such as validity and truth can be checked along the inference dimension of GNOWSYS. Based on the rules, and axioms it is possible to deduce consequences using deductive inference. It is also possible to add ampliative (induction and abduction) and analogical inference engines to the system. This module is not part of the core system but any existing inference engines can be employed using the communication interface of GNOWSYS.

Methodology

In the methodology adopted by us, we have followed the KR approach to organize biological knowledge and create a knowledge base. The steps followed are: to organize concepts on the basis of their cognitive function; to assign valid and authentic semantic relations to the concepts; to analyze the knowledge-base based on the usage of different kinds of semantic relations applied; to compare the novice's knowledge structure with that of an experts knowledge structure; to restructure (reorganize) to align the novice's knowledge structure with the experts knowledge structure; and finally to develop a minimal set of relation types (knowledge organizers) for representing the entire domain of biology. Graphical representation such as concept maps, concept graphs can be generated based on the knowledge base. We would like to point out that at this stage, we have managed to fulfill the first two objectives which are presented in this article and the latter ones are part of our ongoing research project[11].

Organizing concepts on the basis of their cognitive function

Knowledge organizers consists of (1) concepts (ObjectTypes) and the types of concepts (MetaTypes) used in knowledge (2) types of relations used to relate the concepts (RelationTypes) and (3) logical connectors and quantifiers used to express the knowledge. We start with organizing the body of knowledge into concepts and relations (monadic predicates as attributes, and dyadic predicates as relations). The type of concepts are organized in the MetaType layer. The type concepts (general) are organized in the Type layer and the instances (specific) are organized in the Token layer.

Concepts are first to be organized based on their cognitive function depending upon what they explain. For example, 'mitochondria' refers to a structural part of a cell, and so it is instantiated under MetaType *structural concept*; 'protein synthesis' refers to a process and hence it is instantiated under MetaType *process concept*; 'prophase' which is one of the stages involved during the cell division (mitosis), is instantiated under MetaType *stage concept*. It may be noted that relations established in this step are

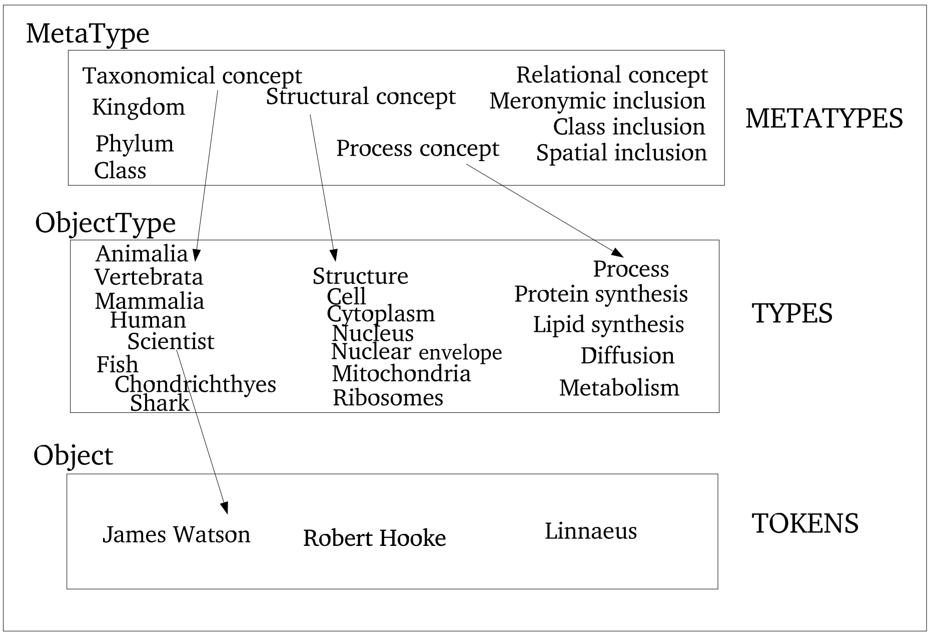

Figure 2: Classifying the concepts based on their cognitive function.

between the concept types and concepts, as shown in figure 2.

Figure 2, shows the organized form of body of knowledge about the structures in the cell. In the metatype layer, the concepts are categorized as structural concept, taxonomical concept, relational concept etc. In the type layer, the structures, process are represented (in the form of an instance of the metatype layer); and the different kinds of structures of the cell (nucleus, mitochondria etc.) or processes of the cell (protein synthesis, lipid synthesis) are organized as the subtypes of the types within type layer. Most of the knowledge is organized in the type layer. The last layer i.e. the token layer represents the individuals i.e. specific instances of the types. Since in science we talk about generals rather than particulars, in a knowledge base of sciences we do not expect many tokens. However, data collected during experimental activities is most about particulars.

Much of the core biological knowledge is contained in physiology, molecular biology, developmental biology, ecology etc. Based on our earlier analysis of biological terms, most biologically significant knowledge of this field is expressed in terms of concepts that describe processes (events), states, or stages, and cycles[20]. It is possible to represent the structures in biological sciences using the different kinds of inclusion relations, however modeling of processes for knowledge representation is quite challenging. The major concerns for knowledge representation, are regarding time, change, sequence, context etc. According to Sandewall, the processes are classified, based on the complexity, as: *discrete or continuous; linear or branching, independent or ramified; immediate or delayed; sequential or concurrent; predictable or surprising; normal or equinormal; flat or hierarchical; timeless or time-bound; forgetful or memory-bound*[19]. In a process, basically a state gets transformed into another state. Such state-transition processes can also be represented using the *Petri nets*. At present, the component classes required to do process modeling are being developed.

Assigning authentic and valid semantic relations to the concepts

Our understanding and expression of knowledge of the world depends on characterizing and establishing relationships between concepts. Structuring of relationships serve as foundations for organizing knowledge. Relations are ubiquitous, and play a central role in our mental and external worlds. "Our mental world—that is, knowledge—is in turn full of representations that correspond in salient ways to the external world"[17]. Merely classifying concepts into different organizational layers is not sufficient. Concepts get meaning on the basis of the semantic relations with other concepts. So, concepts have to be assigned valid and authentic semantic relations.

In figure 3, we present different kinds of semantic relations assigned for the concepts for representing some of the anatomical details of the cell and also to represent some of the taxonomic classification. Semantic relations have been categorized as inclusion (meronymic, class, spatial), attribution, attachment etc[21]. The relation types used in figure 3 are—*includes, part-of, located-on, wound-around, occurs-in, function*, etc. The

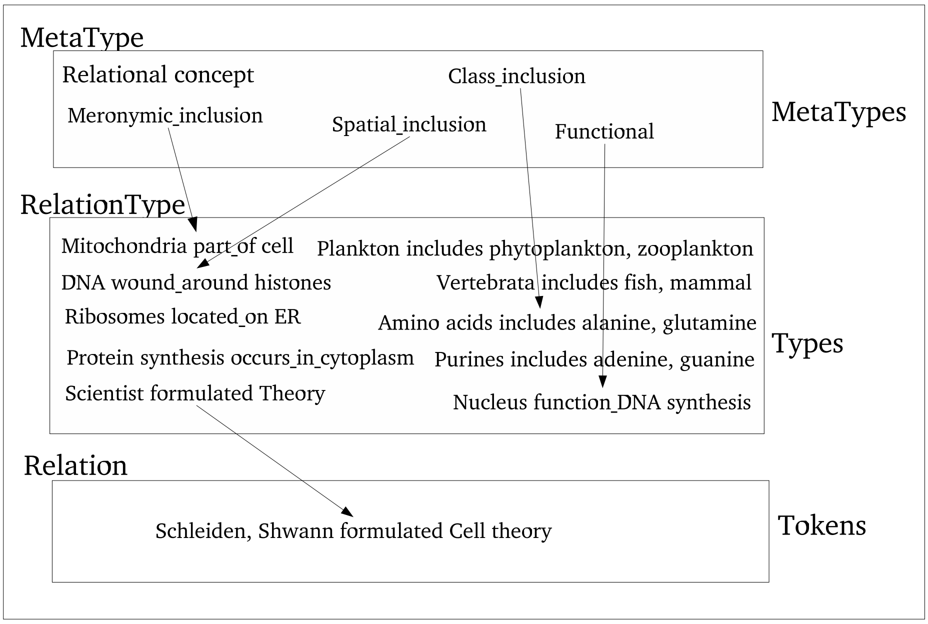

Figure 3: By assigning valid semantic relations between concepts, knowledge base is built.

propositions thus created are as follows—*vertebrata includes fish, mammal*; *plankton includes phytoplankton, zooplankton*; *amino acids includes alanine, glutamine*; *purines includes adenine, guanine* (class inclusion); *mitochondria part-of cell* (meronymic inclusion); *DNA wound-around histones*; *ribosomes located-on endoplasmic reticulum*; *protein synthesis occurs-in cytoplasm* (spatial inclusion); *nucleus function DNA synthesis* (function relation). Currently the support for specifying scope of relations (quantification) and modality of assertions are being developed in GNOWSYS. This feature enables to represent propositions such as: *Some ribosomes located-on endoplasmic reticulum; All DNA wound-around histones in Eukaryotic cell.*

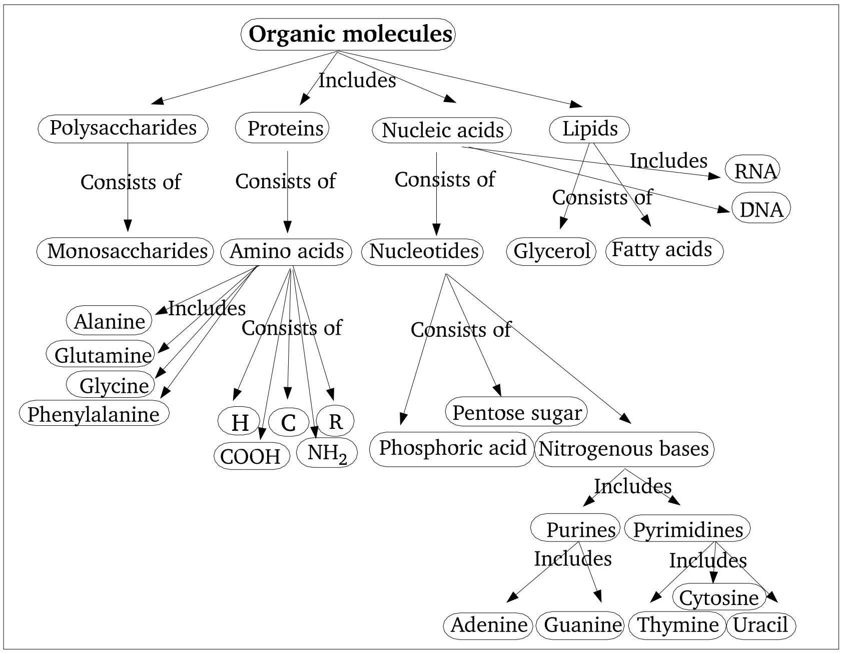

Figure 4: Knowledge organizers for understanding organic molecules. A concept map of biomolecules drawn using principled approach using just 2 relation types (knowledge organizers).

Drawing concept map following the principled/logical approach

A principled or a logical approach is based on the grammar of scientific knowledge. We can draw principled or logical based concept maps (using the KR approach) which are different from the concept maps influenced by Novak.

Figure 4, shows a principled concept map on organic molecules generated from our knowledge base. The concept map depict the minimal set of knowledge organizers i.e. relation types for representing the biomolecules. It is possible to represent some of the knowledge about organic molecules with using just **two relation types** i.e. semantic relations (knowledge organizers) as depicted in the figure. We intend to develop a set of such knowledge organizers for each domain which can be applied in science education. In the domain of KO, various tools exists such as Cyc[12], UMLS[16] etc. which have been built using the logical approach. The principled concept map is very close to the way an expert tries to represent scientific knowledge in terms of unambiguous relations. Parsimony is maintained in the representation of scientific knowledge and the relations used are clear and precise which helps for a better understanding of scientific knowledge after teaching normal concept maps. We propose that if we teach the students to draw concept maps using this principled/logical or KR approach, then restructuring of knowledge can occur in a novice which helps in transforming a novice into an expert, which is the goal of science education. However, for each domain of science an acceptable grammar of science among the practitioners should be arrived after careful discussions.

Conclusion

The study is regarding characterizing and organizing knowledge based on KR using the grammar of scientific knowledge. We introduce the model for representing a small domain of biology. The purpose of this communication is limited to share the approach and assumptions followed. Our methodology sought to classify and organize a small domain of biological knowledge and arrive at a minimal set of knowledge organizers for representing the structural relations in the biological domain. Using these knowledge organizers, we can eliminate ambiguity, maintain parsimony and apply precision to the scientific body of knowledge. At present, we are working on building a process ontology inorder to represent the processes (events), states or stages and cycles of the biological sciences.

We propose that if the students (novices) are trained to characterize and organize knowledge using KR principles i.e. following the grammar of scientific language, then their conceptual framework can be aligned with that of experts' conceptual framework. We think that the principled approach, in place of standard concept mapping, has a direct role to play in science education.

References

[1] David Ausubel, Joseph Novak, and Henry Hanesian. *Cognitive Physchology: A Cognitive View*. Holt, Rinehart and Winston, New York, 1978.

[2] William Brewer and Ala Samarapungavan. Children's theories vs. scienctific theories: Differences in reasoning or differences in knowledge? In Hoffman and Palermo, editors, *Cognition and the Symbolic Processes: Applied and Ecological Perspectives*, pages 209–232. Erlbaum, NJ., 1991.

[3] Susan Carey. Conceptual change and science education. *American Psychologist*, 41(10):1123–1130, 1986.

[4] Kathleen Fisher and Michael Kibby, editors. *Knowledge Acquisition, Organization, and Use in Biology*. Springer-Verlag, Germany, 1996.

[5] Kathleen Fisher, James Wandersee, and David Moody. *Mapping Biology Knowledge*. Kluwer Academic Publishers, The Netherlands, 2000.

[6] Nagarjuna. G. From folklore to science. In *Teaching and Communicating Science: What the History, Philosophy and Sociology of Science can Contribute?*, Leeds, England, 2005. IHPST. To Appear in the Proceedings.

[7] Nagarjuna G. Gnowsys: A system for semantic computing. Website, 2005. `http://www.gnu.org/software/gnowsys/gnowsys_files/cpv2/`.

[8] Nagarjuna G. Gnowsys: Gnowledge networking and organizing system. Website, 2005. `http://www.gnu.org/software/gnowsys/`.

[9] Nagarjuna. G. Layers in the fabric of mind: A cognitive ontogeny. In Jayashree Ramadas and Sugra Chunawala, editors, *Episteme 1: An International Conference to review Research on Science, Technology and Mathematics Education*. HBCSE, HBCSE, India, 2005. To Appear in the Proceedings.

[10] Nagarjuna. G and Meena Kharatmal. Towards a principled approach of concept mapping. Working Draft.

[11] Meena Kharatmal and Nagarjuna G. Knowledge organization of biology education. Website, 2004. `http://zope.hbcse.tifr.res.in:9673/Meena/Biology/`.

[12] Doug Lenat. The cyc knowledge server. Website, 2005. `http://www.cyc.com/cyc/technology/`.

[13] Joel Mintzes, James Wandersee, and Joseph Novak, editors. *Teaching Science for Understanding — A Human Consctructivist View*. Academic Press, USA, 1998.

[14] Joel Mintzes, James Wandersee, and Joseph Novak, editors. *Assessing Science Understanding – A Human Constructivist View*. Academic Press, USA, 2000.

[15] Joseph Novak and D. Bob Gowin. *Learning How to Learn*. Cambridge University Press, UK, 1984.

[16] National Library of Medicine. Unified medical language system (umls). Website, 2005. `http://www.nlm.nih.gov/research/umls/`.

[17] Carol A. Bean Rebecca Green and Sung Hyon Myaeng. In Carol A. Bean Rebecca Green and Sung Hyon Myaeng, editors, *The Semantics of Relationships—An Interdisciplinary Perspective*, chapter Introduction, pages vii–xvi. Kluwer Academic Publishers, London, 2002.

[18] John Sowa. *Conceptual Structures: Information Processing in Mind and Machine*. Addison-Wesley Publishing Company, USA, 1984.

[19] John Sowa. *Knowledge Representation: Logical, Philosophical and Computational Foundations*. Brooks/Cole, USA, 2003.

[20] Sandhya Thulasidas and Nagarjuna. G. A methodology for the analysis of biological knowledge base. In S. C. Agarkar and V. D. Lale, editors, *CASTME-UNESCO-HBCSE International Conference on Science, Technology and Mathematics Education for Human Development*, volume 1, pages 149–155. HBCSE, Mumbai, India, 2001.

[21] Morton Winston, Roger Chaffin, and Douglas Herrman. A taxonomy of part-whole relations. *Cognitive Science*, 11:417–444, 1987.