Ontology-Inferred Phylogeny Reconstruction for Analyzing the Evolutionary Relationships between Species: Ontological Inference versus Cladistics

Arash Shaban-Nejad, Volker Haarslev

Department of Computer Science and Software Engineering, Concordia University, H3G 1M8, Montreal, Canada

Email: {arash_sh, haarslev}@cs.concordia.ca

Abstract— **We propose the use of formal ontological inferencing, rather than cladistics, to reconstruct phylogeny trees and to analyze the evolutionary relationships between species. For this experiment, we focused on the phylogeny of fungi. Lexical chaining technique has been used for incremental population of evolving ontological elements. Also category theory has been employed to provide an underlying formalism for capturing and analyzing the evolutionary behavior of the system.**

Keywords: Ontology, Phylogeny, Fungal Genomics, Ontology population, Category Theory,

I. INTRODUCTION

The major efforts to reorganize taxonomies of species over time can be summarized as the dynamic identification of essential classifying properties for a class and the collection of all beings that share values for these properties into that class [1]. It is commonly believed that all species are descended from a common ancestral gene pool through gradual divergence [2] and form different kingdoms in the tree of life.

In this process of constant evolution, Fungi were promoted from one subclass in the Plant kingdom to a kingdom of their own based on gene mutation. A gene mutation, whether hereditary or new is a permanent change in the DNA sequence that makes up a gene [3]. These changes, which can be insertions, deletions or rearrangements of genetic information happen in relation to time and alter the evolutionary taxonomies of different species. Thus, through several mutations, the fungal classes are promoted, moved, folded, deleted, merged, and renamed as more is discovered about life on Earth. One of the primary goals of taxonomists is to reflect evolutionary history (phylogeny) in the biological classification [4]. Phylogenetic trees demonstrate how a group of species are related to one another. To analyze the evolutionary relationships between groups of organisms for the purpose of constructing family trees, biologists currently use a method called cladistics or "phylogenetic systematics". Through this method, organisms are classified based on their evolutionary relationships; to discover these relationships, primitive and derived attributes should be analyzed [5]. An extensive collection of evidences for importance of systematics and taxonomy (with emphasize on fungi taxonomy) in biological research recently become available, provided by researchers from the British Mycological Society at (http://www.parliament.uk/parliamentary _committees/lords_s_t_select/evidenceselect.cfm).

In summary, cladistics is based on the following assumptions [6]:

1. Any group of organisms is related by their descent from a common ancestor. Therefore, there is a meaningful pattern of relationships between all collections of organisms.
2. The taxonomic trees should be binary, which means that new organisms may come into existence when currently existing species divide into two groups.
3. Changes in attributes occur in lineages over time.

The third statement is the most important rule in cladistics. In fact, only when attributes and characteristics change one can recognize various lineages or groups [6]. Cladistic analysis has proved useful for analyzing evolutionary trees, but it does face several issues, mostly addressed in [5], [7].

In order to overcome some of the issues that affect the cladistic inferencing, we have employed the FungalWeb Ontology [8], a formal ontology empowered by logic as a conceptual backbone to provide a common formal specification for each species in the fungal evolutionary tree. "Lexicon chaining" as a natural language processing (NLP) technique has been proposed for dynamically populating the ontology. To analyze the temporal fungal phylogeny, we also use category theory, which provides an underlying mathematical knowledge representation language. This paper is structured as follows: Sections 2 and 3 discuss about the cladistic technique for studying evolutionary trees and the related issues. In the Section 4, the relations between

ontology, taxonomy and phylogenies are utilized. In Section 5, we introduce our proposed ontology-driven method, which facilitates semi-automatic phylogeny construction for analyzing evolutionary relationships between species. Section 6, discusses about using category theory as the underlying formalism for our framework. Sections 7 and 8 are focused on the evaluation method and the related works respectively.

II. PHYLOGENETIC SYSTEMATICS (CLADISTICS)

The taxonomical classification has a long history in biology; since the time of Darwin (1809–82) and his theory of natural selection [10] there have been debates between two groups of taxonomists [4]:

1. Classical taxonomists working on "Linnaean classification" [11], a system based on a hierarchy of formal ranks (family, genus, etc.) and binomial nomenclature.

2. Cladists working on phylogenetic classification or cladonomy [11], which is a clade-based classification system, without any formal ranks, including the genus, and no binomial nomenclature [11], [12].
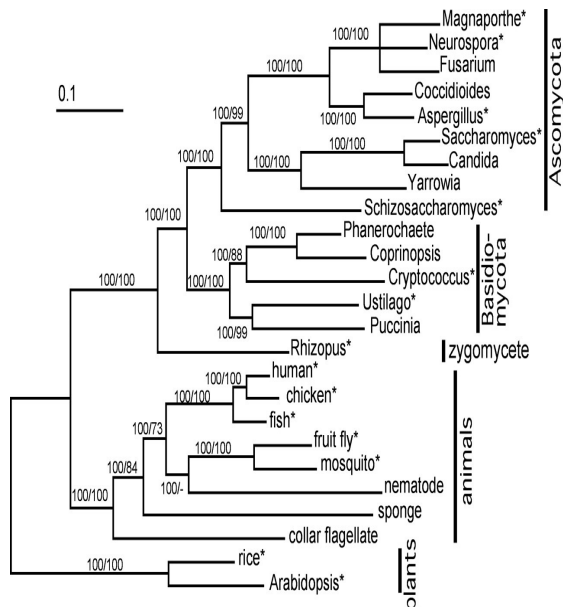

Fig. 1. An example of a phylogenetic tree including different clades of fungi, animal and plants for demonstrating variation in rates of nucleotide substitution (Source: adapted from [46]).

Cladistic approaches are being used to analyze the evolutionary trees based on primitive and derived attributes. Primitive attributes (plesiomorphic) are those attributes of a fungus that are shared by all members of the group. Having "fruiting body" is a primitive attribute for all species of Basidiomycota (a major phyla in the fungi kingdom), which has been inherited from their common ancestor. Primitives are not very helpful for analyzing the reltionship between organisms in a specific group [5].

When we try to construct a family tree for all Basidiomycotas, it is not helpful to note that they all have fruiting bodies, and it does not help us in determining the relationships between different species. Derived attributes (apomorphic) are advanced feature that only appear in a number of members [5]. In fact, the derived attributes are crucial to construct evolutionary relationships. For example, the shared derived attribute that defines the Ascomycota is the ascus [13]. Nuclear fusion and meiosis occur inside the ascus where one round of mitosis follows meiosis to leave 8 nuclei, and 8 ascospores [13], [14]. Accordingly, Fungi can be divided into two biological groups: without ascus and with ascus. The intersection of these two groups (a node) can be represented in an evolutionary diagram (cladogram) as a point at which a new species (with ascus fungi) evolved [5]. Having ascus is a synapomorphy (a derived attribute shared by two or more taxa) of the Ascomycetes group. In cladistic method synapomorphies are used to construct phylogenies. A synapomorphy of one group might be primitive for another group. By analyzing sufficient attributes cladistics aims to generate a family tree where either all members are descended from a single, common ancestor (monophyletic) or from several common ancestors (polyphyletic) [5]. If the group includes some, but not all, of the descendants of a single common ancestor, it is called paraphyletic [15].
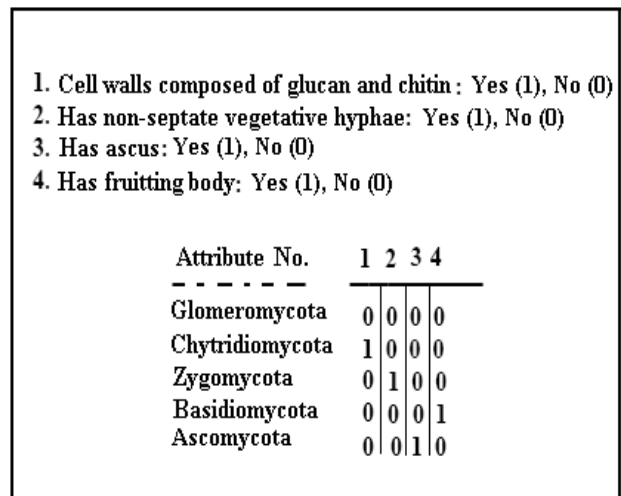

Fig. 2. An example of a sample data matrix for analyzing major fungi clades (Ascomycota, Basidiomycota, Zygomycota, Chytridiomycota and Glomeromycota.

Cladistic analysis is currently performed using various software applications such as PHYLIP (Phylogeny Inference Package) [51], PAUP [16] and MacClade [17].

A data matrix similar to the one demonstrated by Figure 2 provides the input for cladistic analysis. This matrix simply summarizes the answers to questions such as: does a fungus have a set of attributes, or not? The answers are short and simple ([yes, no] or [1, 0]). The more species and the more

attributes one puts in an analysis, the more likely it gets close to the accurate family tree [5].

III. ISSUES IN CLADISTIC ANALYSIS

There are some issues in cladistic analysis [5]:

I. Convergent evolution: If one defines having a fruiting body as an attribute of fungi basidiomycota, and considering that many plants have also fruiting bodies, should basidiomycota be considered closer relatives of plants than of the ascomycota fungi? The answer is negative. In fact, basidiomycota and ascomycota have a number of shared derived attributes that closely link them. Convergent evolution produces homoplasies. A homoplasy [18, 19] can be defined as: "a resemblance between taxa that can be ascribed to processes other than descent from a common ancestor and which implies phylogenetic relationships that conflict with the best estimate of phylogeny for the taxa" [20]. By providing and analyzing as many different attributes as possible this problem can be reduced [5].

II. Reversals can cause problems: As an example, whales unlike all the mammals do not have fur, because the fur of their mammalian ancestors has been lost in an aquatic environment [5], [21].

III. Considering fossils with missing parts: In this case, the attributes associated with those missing parts are represented by question marks and ignored when generating the cladogram.

IV. FORMAL ONTOLOGY, TAXONOMY AND PHYLOGENETIC ANALYSIS

Taxonomy in knowledge representation is considered as a collection of terms or entities organized in a hierarchical structure (implying parent-child relationships). The new trend in analyzing taxonomical relationships is emerging to use ontology, as defined by Gruber [22] "specification of conceptualization", to provide an underlying discipline of sharing knowledge by defining concepts, properties, and axioms. Ontologies in the context of semantic web consist of "taxonomies and a set of inference rule" [47]. There may be more than one taxonomy for an ontology in a domain of interest, based on the granularity and the chosen subsets of ontological characteristics.

Ontologies in the real world evolve over time as we fix errors, reclassify the taxonomy, and add or remove concepts, attributes, relations, and instances. Consistently modifying and adjusting the hierarchical structure of ontologies in response to changing data or requirements can provide new insight for studying evolutionary changes (or mutations in evolutionary phylogenies) in biological taxonomies occuring over time. Ontologies follow the open world assumption, which asserts that the captured knowledge is always incomplete, therefore if something cannot be inferred from what is defined in the knowledgebase, it is not necessarily false. The open world assumption is especially important when we represent knowledge with a dynamic system, which is gradually improved as we discover new facts. In cases such as the real world phylogeny analysis our knowledge is always incomplete and the facts described by the system can never be fully known. Due to the evolutionary nature of cladistics, it is possible to study the way in which attributes change (the direction in which attributes change, and the relative frequency of the change) over time within groups [23] in an ontological framework. In order to study various changes in ontological inferred phylogenetic tree one can focus on ontology evolution and change management techniques. "Ontology evolution" aims to maintain the dynamic structure of ontologies and controlled vocabularies, to preserve the validity and consistency of ontological knowledge.

Analyzing the fungal taxonomy within the FungalWeb framework facilitates ontological inferencing - which provides a valuable source of information for clarifying the explanations of complex evolutionary scenarios for fungi species - rather than cladistics inferencing. The ontology inferencing allows us looking at the diversity of the species within different groups by comparing the descendants of an ancestor to find out the patterns of origin and extinction. It also empowers biologists to examine different hypotheses about adaptation [10], [23]. Currently, there is a need for a comprehensive methodology to describe how chronological alterations in ecological and environmental conditions [24] have formed the adaptive evolution of fungal clades.

V. METHODS, RESULTS AND APPLICATION SCENARIOS

By changing the knowledge, ontologies need to be incrementally updated to provide valid information for the human/agent learner. In our approach, we have used the Lexical chaining method to (semi-) automatically construct and populate the FungalWeb ontology by extracting relevant terms and relations from a structured or unstructured text corpus or other types of data. The Lexical chaining algorithm [25] reads a text corpus and places words in a related chain based on semantic similarity, using a set of reference dictionaries such as WordNet 3.0 (http://wordnet.princeton. edu/), Integrated Taxonomic Information System (ITIS) (http://www.itis.gov/) and TreeBase (a database of phylogenetic knowledge) (http://www.treebase.org/treebase/). Then, using an agent-based framework [26], the related ontologies – which provide the underlying knowledge for the learner agent – can be dynamically populated and validated using description logics [27] reasoner (e.g. RACER) [28] (Figure 3).

If some species have similar properties and genomes, it is very likely that they evolved from a common ancestor. The similarity of genomes is computationally measured based on

the number and likelihood of different mutations (insertion, deletion, duplication or substitution of base pairs) [29].
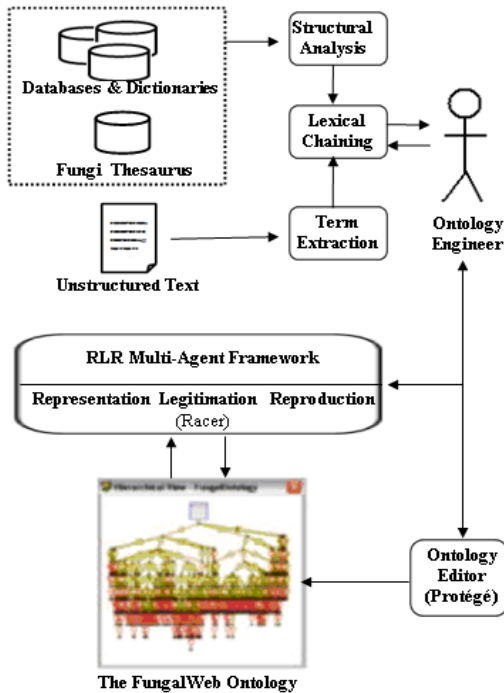

Fig. 3. Framework for ontology learning and population

We have used the FungalWeb Ontology to determine the taxonomic provenance [8] for fungal species, in order to study the evolutionary relationships based on logical and ontological inferencing.

Identifying taxonomic provenance is crucial within the Gene Discovery process. For instance, an enzymologist may want to know which fungi are known to produce the enzyme 2-deoxyglucose-6-phosphatase (EC# 3.1.3.68), and the common lineage that these organisms share. By querying the FungalWeb Ontology the enzymologist can find the related fungal species: Pichia stipitis and Saccharomyces cerevisiae.

Identifying the common lineage between the found organisms requires identifying the highest taxonomic group that unites all species known to produce the enzyme of interest, akin to finding a common ancestor [8]. Within the FungalWeb Ontology, a fungal taxonomy is represented in a deep hierarchy of taxonomic units/concepts. The defined key properties between "fungi" and "enzyme" allow for the identification of species found to produce 2-deoxyglucose-6-phosphatase.

One can identify the common lineage for these fungal species by using the description logic reasoner, RACER, via the command *instance types,* which retrieves the concepts that instantiate each fungal species individual. A simple example of such queries is shown in Query 1. The common lineage of "2-deoxyglucose-6-phosphatase"-producing fungi, is a family of yeast in the order *Saccharomycetales* called

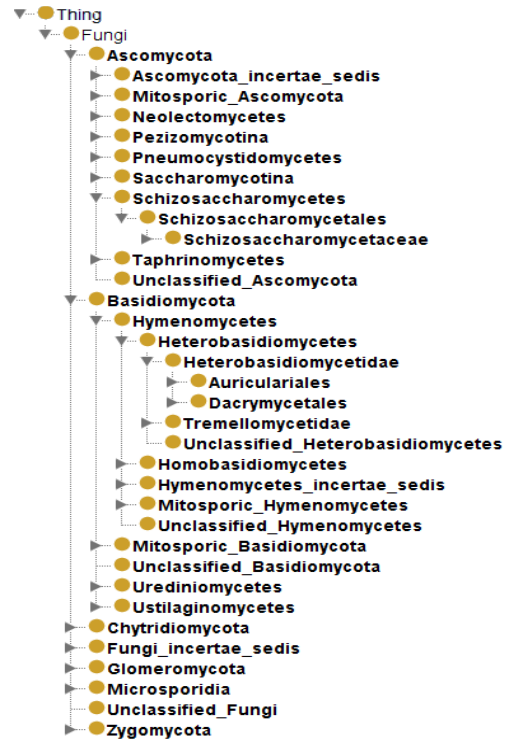*Saccharomycetaceae*, known for its reproduction by budding and ferment carbohydrates (WordNet definition).

Fig. 4. Domain model of fungal taxonomy

Query 1: This query uses RACER command "Instance types" to retrieve results for all fungi that produce the enzyme 2-deoxyglucose-6-phosphatase (EC# 3.1.3.68) as well as their ancestors. The common subset identifies the common lineage between the species:

    <<:?X :http://a.com/ontology#Fungi:>
    <<:?X :http://a.com/ontology#Ascomycota:>
    <<:?X :http://a.com/ontology#Saccharomycotina:>
    <<:?X :http://a.com/ontology#Saccharomycetes:>
    <<:?X :http://a.com/ontology#Saccharomycetales:>
    <<:?X :http://a.com/ontology#Saccharomycetaceae:>

We are currently working on different aspects of managing the evolving structure of the FungalWeb Ontology – both syntactic and semantic changes – that can be used to automate the phylogeny tree reconstruction and define a meaningful pattern of relationships between the species.

VI. CATEGORY THEORY AS UNDERLYING FORMALISM

We have also employed [9], [26] category theory [31] with its set of objects and morphism (which are comparable with sets of ontological concepts and relationship arrows) as a mathematical vehicle to represent, analyze, and track the changes in the evolutionary tree. Category theory facilitates the analysis of the process of structural relationships and structural change in living and evolving systems. The

abstractness of category theory facilitates the description of domain independent expressions. It also can be used for composition mechanisms, to address scalability issues [32].

By using Functor (morphisms in the category of all small categories) we describe the set of state space (set of all possible states for a given state variable set) for a class as a cross product of attribute domains and the operations of a class as transitions between states for ontological elements indexed by time.

To predict the direction of a change and its consequences in our framework we chose the coupling measurement
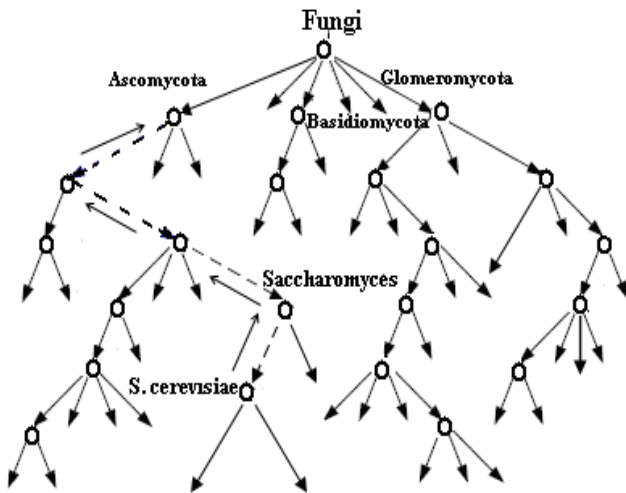

Fig. 5. The categorical representation of ontology inferred phylogeny for yeast Saccharomyces cerevisiae which depicts the transition between various evolutionary states.

approach that we introduced in [9].

We also define some operations for each ontological class. As demonstrated in Figure 2, position of numbers in the data matrix indicates change from state 0 (primitive condition) to state 1 (derived or advanced state). In ontology, a concept or an instance can transit from one state to another based on its behavior in response to a change. An event can be formally modeled as an ordered pair $E = <St_1, St_2>$ [37]. St_1 is the start state at time t1 and St_2 is the end state at time t2. St_1 and St_2 are not necessarily distinct; they might refer to the same state. The categorical representation of ontologies (Fig. 5) along with other formalisms such as description logics, enables us to capture the full semantics of evolving hierarchies (See [9] for more information).

As demonstrated in Figure 5, category theory is capable of solving problems related to reverse analysis (mentioned in cladistics method) through *recursive domain equations* [33]. In order to analyze the bifurcating pattern of cladogenesis, which states that "new organisms may come to exist when currently existing species divide into exactly two groups" [6], we have used two categorical constructors: pushouts and pullbacks. The *pushout* for two morphisms f:A→B and g:A→C is an object D, and two morphisms i_1:B→D and

i_2:C→D, such that the square commutes (Fig 6.a). D is an initial object in the full subcategory of all such candidates D′ (i.e., for all objects D′ with morphisms j1 and j2, there is a unique morphism from D to D′). The *pullback* (also known as "Cartesian square") for two morphisms f:A→C and g:B→C is an object D, and two morphisms i_1:D→A and i_2:D→B, such that the square commutes. Here D is the terminal object in the full subcategory of all such candidates D′ [41] (Fig 6.b).

Using *pushout* and *pullback* as it is shown in [34], [35] facilitates studying merging, composition and decomposition of evolutionary taxonomical structure in the categorical framework.
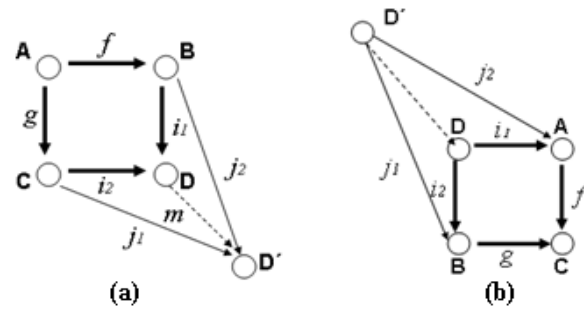

Fig. 6. (a) Pushout, (b) Pullback

Placing an organism in a phylogeny tree and associating a set of roles based on its evolutionary characteristics may sometimes lead to redundancy in the taxonomy. One of the major issues in phylogeny analysis is finding and identifying equivalent classes and relationships. Category theory enables us to deal with the problem of logical equality [48] by using a categorical constructor called isomorphism. A morphism f: A→B of category C is called an isomorphism iff there is a morphism g:B→A of category C such that composition fg: A→A (or id$_A$) and gf: B→B (or id$_B$). Bijections in the category of sets are example of isomorphism (See [48] for details).

VII. EVALUATION AND VERIFICATION

For the evaluation of the quality of the populated ontology through lexical chaining some criteria such as reiteration, density and length of the chains [50] can be considered. The legitimation phase in our agent-based framework [26] is also responsible to assess the impact of a potential change before the change is actually made. Logical legitimation are obtained by a reasoning agent, which is a software agent that controls and verifies the logical validity of a system, revealing inconsistencies, misclassifications, hidden dependencies and redundancies. It automatically notifies users or other agents when new information about the system becomes available. We use RACER as a description logic reasoner agent, along with other semi-automated reasoning

system for basic category-theoretic reasoning based on a first-order sequent calculus [36]. It captures the basic categorical constructors and provides services to check consistency, semantic coherency, and inferencing [36].

VIII. RELATED WORK

In the last decade, several efforts [38], [39], [40] have been reported in the pursuit of comprehensive frameworks for maintaining hierarchical structures and evolving ontologies. Since existing knowledge representation languages, including well-established description logics, can not guarantee the computability of highly expressive time-dependent ontologies, current efforts are mostly focused on time-independent ontological models. However, the real biomedical ontologies exist in time and space. In our approach, category theory with its rich set of constructors can be considered a complementary knowledge representation language to capture and represent the full semantics of evolving phylogenetic trees.

Rosen (1958) [43] was among the first who proposed the use of category theory in biology, in the frame of a ''relational biology''. In [42] category theory has been used for analyzing tree transformations with considering relabeling (only tree node's labels are changed), and restructuring the tree (when the tree structure is not preserved, either through rebranching or relayering). Also, category theory has been proposed by [34], [35] to study ontology alignment and merging.

IX. DISCUSSION

A clear identification of evolutionary relationships of species provides important information for understanding and characterization of the genetic diversity. Evolutionary speaking, mutations, either partial or complex, can cause transition, transversion, inversion, deletion, insertion, duplication, translocation and elongation [45] in the genetic structure of species, which alter the related phylogenies. Due to the several problems in the cladistics analysis method we have proposed our approach for constructing a formal ontology-driven fungi phylogeny. Logically described ontologies provide facilities to reconstruct and manage the evolving structure of phylogenetic trees. Based of the open world assumption in ontologies, the inferred phylogenies are always seen as evolving source of knowledge, which provide open-ended answers to the posed queries. Also, using formal reasoners assists revealing hidden dependencies as well as redundancy and misclassification in the inferred hierarchy. Using lexical chaining facilitates hierarchical organization of sequentially described terminologies to dynamically populate the ontology. Ontological framework also helps sharing common concepts between different applications and reusing each inferred phylogeny in an integrated system. Our categorical approach draws its inspiration from hierarchical

systems of categories where ontologies can be seen as an interconnected hierarchy of theories as a sub-category of a category of theories expressed in a formal logic [52]. Category theory has a rich structure with a precise language and convenient symbolism for visualization. Using categories with its intuitive, yet efficient constructors enables us to formalize the temporal structure of an evolving system.

Some of the challenges that we faced in applying our approach are as following: In the task of employing lexical chaining algorithm we had the problem of non-cohesive [49] text corpuses which dramatically reduce the efficiency of our approach. Therefore we decided to start with the assumption that the target text is cohesive. Another problem is back to the ontological completeness. Although the use of ontology inferred phylogeny is a very useful way forward, its success highly depends on taxonomic expertise and the availability of rich consistent collections of defined concepts for accurate and precise inferencing. For future research we plan to extend the use of category theory with colored Petri-net to enhance the taxonomical visualization and improve querying and tracking capabilities.

REFERENCES

[1] H.S. Pinto, S. Staab, C. Tempich, "DILIGENT: Towards a fine-grained methodology for Distributed, Loosely-controlled and evolving Engineering of oNTologies," *in Proc. 16th Eureopean Conference on Artificial Intelligence, (ECAI'04)*, Valencia, 2004, pp. 393–397.

[2] D. J. Futuyma, *Evolution.* Sunderland, Massachusetts: Sinauer Associates, Inc. 2005, ISBN 0-87893-187-2.

[3] Mutations and Health from Genetics Handbook, Genetic Home Reference. Available: http://ghr.nlm.nih.gov/ , June 20, 2008.

[4] Taxonomy, Classification, and the Debate about Cladistics, From an appendix in Shinners & Mahler's Illustrated Flora of North Central Texas; 1999, BRIT & Austin College. Available: http://artemis.austincollege.edu/acad/bio/gdiggs/taxonomy.html

[5] L.M. Clos, "What is cladistics?," *Fossil News, Journal of Avocational Paleontology*, 1996. Available: http://www.fossilnews.com/1996/cladistics.html

[6] Phylogenetic systematics, a.k.a. evolutionary trees. The centre for understanding evolution, Berkeley University. Available: http://www.ucmp.berkeley.edu/clad/clad1.html

[7] H. Robinson, "A Key to the Common Errors of Cladistics," *Taxon*, vol. 35(2), 1986, pp. 309–311.

[8] C. J. O. Baker, A. Shaban-Nejad, X. Su, V. Haarslev, and G. Butler, "Semantic web infrastructure for fungal biotechnologists," *Journal of Web Semantic,* vol. 4(3), 2006, pp. 168–180.

[9] A. Shaban-Nejad, and V. Haarslev, "Categorical Representation of Evolving Structure of an Ontology for Clinical Fungus," in *Proc. of 11th Conference on Artificial Intelligence in Medicine (AIME'07)*, Amsterdam, Springer, LNCS, Vol. 4594, 2007, pp. 277– 286.

[10] Why Do Biologists Need Cladistics? Available: http://www.Ucmp.berkeley.edu/clad/clad5.html

[11] R. K. Brummitt, "Taxonomy versus cladonomy, a fundamental controversy in biological systematic," *Taxon,* vol. 46, 1997, pp. 723–734.

[12] K. de Queiroz, and J. Gauthier, "Phylogenetic taxonomy," *Ann. Rev. Ecol. Syst.*, vol. 23, 1992, pp. 449–480.

[13] C. G. Wu, and J. W. Kimbrough, "Ultrastructural studies of ascosporogenesis in Ascobolus immerses," *Mycologia*, vol. 84, 1992, pp. 459–466.

[14] J. W. Taylor, J. Spatafora, and M. Berbee, "Ascomycota. Sac Fungi," Version 09 Oct. 2006. Avialable:

http://tolweb.org/Ascomycota/20521 /2006.10.09) in The Tree of Life Web: Project: http://tolweb.org/

[15] The Nature Journal Glossary. Available: http://www. nature.com/nrg/ journal/v3/n11/glossary/nrg929_glossary.html

[16] D. Swofford, "PAUP (V. 4.0) A tool for inferering and interprtting phylogenetic trees," Available: http://paup.csit.fsu.edu/

[17] D. R. Maddison, and W. P. Maddison, "MacClade: a computer program for phylogenetic analysis," published by Sinauer Associates, Avilable: http://macclade.org/index.html

[18] G. G. Simpson, "Principles of Animal Taxonomy," New York: Columbia University Press, 1961.

[19] D. B. Wake, "Homoplasy: the result of natural selection, or evidence of design limitations?," *Am Nat,* vol.138, 1991, pp. 543–567.

[20] M. Collard, and B. Wood, "Homoplasy and the early hominid masticatory system: inferences from analyses of extant hominoids and papionins," *J. of Human Evolution*, vol. 41(3), 2001, pp. 167–194.

[21] Mammal Encyclopaedia Article, Available: http://www.naturalresearch.org/ Mammal/encyclopedia.htm

[22] T. R. Gruber, "A translation approach to portable ontologies," *Knowledge Acquisition*, vol. 5(2), 1993, pp. 199–220.

[23] R. H. Zander, "On the Present Revolution," Buffalo Museum of Science Website, June 2002, Available: http://www.mobot.org/plantscience/resbot/Phil/Revolution.htm

[24] N. M. Andersen, "Cladistic Inference and Evolutioanry Scenarios: Locomotory Structure, Function, and Performance in Water Striders," *Cladistics*, vol. 11(3), 1995, pp. 279–295.

[25] G. Hirst, and D. St-Onge, "Lexical chains as representations of context for the detection and correction of malapropisms," In: Fellbaum, C. (ed.): WordNet: An Electronic Lexical Database. MIT Press, Cambridge, MA, 1998.

[26] A. Shaban-Nejad, V. Haarslev, "Incremental Biomedical Ontology Change Management through Learning Agents," in *Proc. of 2nd KES Intl. Symposium on Agent & Multi-Agent Systems (KES-AMSTA'08)*, Korea, LNCS, Vol.4953, Springer, pp.526–535

[27] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. Patel-Schneider (eds.), "The Description Logic Handbook: Theory, Implementation and Applications," Cambridge University Press, March 2003.

[28] V. Haarslev , and R. Möller, "RACER System Description," in *Proc. of the 1st International Joint Conference on Automated Reasoning (IJCAR 2001)*, Siena, Italy, June 18–23, 2001, p.701–706.

[29] University of Virginia, CS201J Course material, Fall 2002, Available: http://www.cs.virginia.edu/cs201j-fall2002/ problem-sets/ps4/

[30] C. H. Johnson, M.G. Klotz, J. L. York, V. Kruft, and J. E. McEwen, "Redundancy, phylogeny and differential expression of Histoplasma capsulatum catalases" *Microbiology,* vol. 148, 2002, pp. 1129–1142.

[31] S. MacLane, "Categories for the Working Mathematician," 1971 (corrected 1994) Springer.

[32] T. Mens, "A Formal Foundation for Object-Oriented Software Evolution," in *proc. of IEEE Intl. Conference on Software Maintenance (ICSM'01)*, Florence, Italy, 2001, pp. 549–552.

[33] M. B. Smyth, and G. D. Plotkin, "The Category-Theoretic Solution of Recursive Domain Equations," *SIAM J. on Computing (SICOMP)*, vol. 11(4), 1982, pp. 761–783.

[34] P. Hitzler, M. Krötzsch, M. Ehrig, and Y. Sure, "What Is Ontology Merging? – A Category-Theoretical Perspective Using Pushouts," in *Proc. of the 1st Intl. Workshop on Contexts and Ontologies (C&0) at the AAAI-05*, Pittsburgh, Pennsylvania, July 2005, pp. 104–107.

[35] A. Zimmermann, M. Krötzsch, J. Euzenat, and P. Hitzler, "Formalizing Ontology Alignment and its Operations with Category Theory," in *proc. of Intl. Conference on Formal Ontology in Information Systems (FOIS'06),* Baltimore, Maryland, Nov. 9-11, 2006, pp. 277–288.

[36] D. Kozen, C. Kreitz, E. Richter, "Automating Proofs in Category Theory," in *Proc. of the 3rd International Joint Conference on Automated Reasoning (IJCAR'06),* LNCS, vol.4130, Springer, 2006, pp. 392–407.

[37] A. EC. Ehresmann, and J. P. Vanbremeersch, "The Memory Evolutive Systems as a Model of Rosen's Organism-(Metabolic, Replication) Systems," *Springer,* vol.16, pp. 137–154.

[38] D. E. Oliver, Y. Shahar, E. H. Shortliffe, and M. A. Musen "Representation of change in controlled medical Terminologies," *J. of AI in Medicine,* vol. 15, 1999, pp. 53–76.

[39] N. F. Noy, A. Chugh, W. Liu, and M. Musen, "A Framework for Ontology Evolution in Collaborative Environments," *in the Proc. of 5th Intl. Semantic Web Conference (ISWC'06)*, LNCS, Vol. 4273, Springer, Athens, GA, Nov 5-9,2006, pp. 544–558.

[40] M. C. A. Klein, and N. F. Noy, "A Component-Based Framework for Ontology Evolution," in IJCAI-03, CEUR-WS, vol.71, 2003.

[41] S. Easterbrook, Department of Computer Science, University of Toronto, "Category theory for beginners," Tutorial given at ASE'98, Oct 1998. Availabe: http://www.cs.toronto.edu/~sme/presentations/cat101.pdf

[42] I. Hasuo, B. Jacobs, and T. Uustalu, "Categorical views on computations on trees (Extended Abstract)," In *Proc. of 34th Int. Coll. on Automata, Languages and Programming, (ICALP'07)*, LNCS, vol. 4596, *Wroclaw,* Springer-Verlag, 2007, pp. 619–630.

[43] R. Rosen, "The Representation of Biological Systems from the Standpoint of the Theory of Categories," *Bulletin of Mathematical Biophysics,* vol. 20, 1958, pp. 245–260.

[44] A. I. MacFarlane: Dynamic structure theory, "A structural approach to social and biological systems," *Bulletin of Mathematical Biology*," vol. 43(5), 1981, pp. 579–591.

[45] EMBL-EBI: Types of mutation events and their discreet affects on RNA and polypeptide. Available: http://www.ebi.ac.uk/mutations/recommendations/mutevent.html

[46] J. W. Taylor, and M. L. Berbee, "Dating divergences in the Fungal Tree of Life: review and new analyses" *Mycologia,* vol. 98, 2006, pp. 838–849.

[47] T. Berners-Lee, J. Hendler, and O. Lassila, "The semantic web," *Scientific American,* 2001, pp. 30–37.

[48] B. Mazur, "When is one thing equal to some other thing?," June 12, 2007, Available: http://www.math.harvard.edu/~mazur/preprints/when_is_one.pdf

[49] R. Barzilay, and M. Elhadad, "Using lexical chains for text summarization," In: I. Mani, M. T. Maybury (eds.) Advances in automatic text summarization, Cambridge, MA, The MIT Press, 1999, pp. 111–121

[50] J. Morris, and G. Hirst, "Lexical cohesion computed by thesaural relations as an indicator of the structure of text," *J. Computational Linguistics,* vol. 17(1), March 1991, pp. 21–45.

[51] J. Felsenstein, "PHYLIP (Phylogeny Inference Package) version 3.6", Distributed by the author. Department of Genome Sciences, University of Washington, Seattle, 2005.

[52] M. J. Healy, and T. P. Caudell, "Ontologies and Worlds in Category Theory: Implications for Neural Systems", *Axiomathes Journal,* Springer, vol. 16, 2006, pp. 165–214.

[53] E. May, and W. J. Bock, "Classifications and other ordering systems," *J. Zool. Syst. Evol. Research*, vol. 40(4), 2002, pp-169–194.