Managing Conceptual Revisions in a Temporal Fungal Taxonomy

Arash Shaban-Nejad, Volker Haarslev
Department of Computer Science and Software Engineering, Concordia University,
Montreal, Quebec H3G 1M8, Canada
{arash_sh, haarslev} @cs.concordia.ca

Abstract

Representing and Reasoning about time and change is one of the primary issues in the area of Artificial Intelligence (AI) and Knowledge Representation (KR). Despite the importance of time and change in ontology change management process, current ontology researches are based on time- independent models. One of the reasons is that considering time in ontologies can increase the complexity and a comprehensive and very expressive ontology language is needed to represent it. In this paper we propose using state space model along with category theory as a mathematical notation, which is independent of a specific choice of ontology language and any particular implementation to analyze and represent temporal ontological models in the domain of biomedical applications.

1. Introduction

Using ontology that aims to provide a common vocabulary to capture and represent useful knowledge in a domain of interest is a new trend in AI to manage the inherent complexity of large integrated biomedical systems. We have implemented the FungalWeb [1] Ontology which is a formal bio-ontology in the domain of fungal genomics with a large number of instances. Our goal has been to build a semantic web prototype in the domain of fungal enzymology to represent the fungal genomics knowledge across the Web and make it computationally accessible and semantically understandable for both human and computers. We are now reaching a stage where we wish to develop a change management strategy to update ontological knowledge representations. Ontologies such as living organisms are evolving over the time in order to fix the errors, reclassifying the taxonomy, adding/removing concepts, attributes, relations and instances. However every change in ontologies should be studied in relation with time, but the role of time is not taken into account in current ontology evolution studies. One of the most fundamental questions in our research is: how to represent changes in the conceptualization in a time-dependent ontology? In order to address certain aspects of representation of changes in an ontology driven application in the biomedical domain, in this paper we propose a method based on discrete state model and category theory as a mathematical notation, which is independent of a specific choice of ontology language and any particular implementation.

2. Evolving fungal taxonomy

Fungi are one of the widely used species in industrial, medical, food and biotechnological applications. They are also related to many human, animal and plant diseases, food spoilage and toxigenesis. It is estimated that about 10% of 1.5 million fungal species are known [2] and only small portion of them have an identified usage. The Fungal taxonomy is not stable. Most of the alterations are changes in names and taxonomic structure and relationships.

- **Name changes in Fungal Taxonomy**: Fungal names reflect the data about organisms, and thus, as our understanding of the relationships among taxa increases, names will be forced to change as they no longer convey the correct information to the user [3]. Most fungi names are currently based on the phenotype (visible characteristics of organism) and their appearance. These name changes may cause confusion and affect the validity of different queries. An example about eyespot disease of cereals and issues related to naming its associated fungi is actually represented at [4]. The morphological conceptualization of fungi is not sufficient, and no longer will work because all names based only on morphology have to be re-evaluated. In addition, the phylogenetic based conceptualization also has its own limitations, as sometimes the decision of where to draw the line between different species is not easy to make [3]. To manage this process of continuous change one needs to refer to the nature of ontological structure, where names in taxonomy are only meaningful and valuable once linked to descriptive datasets which were extracted and managed from various databases and literatures in an integrated environment. The incorporation of DNA data is also needed to ensure stability in names and reliable species recognition. Currently only around 1.1% of the estimated 1.5 million species on Earth are represented by DNA sequence data, thus it seems that a very low percentage of the already discovered fungal species are in fact being preserved from the change [6].

- **Changes in Taxonomic Structure:** By advancing in molecular biology one can expect changes in taxonomical structure and relationships. For example by studying some molecular, morphological and ecological characteristics *Glomeromycota* was discovered in 2001 [8] as a new fungal phylum. Another example is the sedge parasite *Kriegeria eriophori* which has never been satisfactorily classified. Recently Ribosomal RNA gene sequences and nucleus-associated ultrastructural characters were analyzed separately and combined to define the new subclass Microbotryomycetidae [9].

3. Time, identity and discrete state-model

Things change in relation with time (Kant 1781). However every change in ontologies should be studied in relation to time, but the role of time is not taken into account in ontology evolution studies. One of the reasons is that considering time in ontologies can increase the complexity and we need a comprehensive and very expressive ontology language to represent it. If a changing concept has various incompatible attributes, then a contradiction might be caused in the updated ontology, because there is no guarantee that all attributes will be changed coherently through time. For example in the FungalWeb ontology, fungal taxonomy which is built based on morphology includes fungi that are not clearly classified in one of the fungi categories (Unclassified Fungi). What happens when for example our classificatory knowledge of one of this unclassified species such as ectomycorrhiza_of_Salix_reinii improves over time? Under which category and based on what attributes it should be reclassified? What happens if one decides to put ectomycorrhiza_of_Salix_reinii under the concept "Basidiomycota" based on some molecular properties? This species does not have fruiting body while one of the essential properties of Basidiomycota is to have a fruiting body. Should the essential property be dropped? In this case how we can be sure that other Basidiomycota species which have been already classified based on their body shape are in the correct position? How can we expect to keep those things persistent and consistent over time? Despite many debates about the nature of "time", in order to conceptualize a thing within a formal ontology, one has to consider there are some essential properties, and without those properties that thing does not exist. The essential attributes form the identity of the substance of a thing. These attributes might be used to express quality or quantity in one

concept. In order to say two concepts C1 and C2 are identical at two different time points t_1 and t_2, they should have common essential properties. But if the properties at different time points are incompatible, then a contradiction and inconsistency follows [5]. So to compare things at different time points we are using time as an index. We propose the following solution: Representing conceptualization of things when they are indexed by times (things-at-a-time), for example from FungalWeb Ontology: "*enzyme* has_pH_optimum at *t*" is rendered as "*enzyme*-at-*t* has_ pH_optimum".

In order to represent different states of our conceptualization we use a categorical discrete state-model, which describes the states and events in the ontological structure using a diagrammatical notion. The discrete state-model can be specified by a state space (all potential states), a set of initial states and a next state function [7].

4. Category theory and ontologies

Category theory is a recent domain of mathematics, introduced and formulated in 1945 [10]. A formal model of objects based on "category theory" is introduced in [11] which allows one to analyze both static and dynamic properties of a design in applications in different areas. Category theory is closely connected with computation and logic [12] which allows an ontology engineer to implement different states of design models to represent the reality. Categorical notations consist of diagrams with arrows. Each arrow *f: X$\rightarrow$Y* represents a function. A Category C includes:

- A class of objects and a A class of morphisms ("arrows") and for each morphism *f* there exists one object as the domain of *f* and another as the codomain (Figure 7.1.a).
- For each object, A, an identity morphism which has domain A and codomain A ("ID$_A$"). (Figure 7.1.b)
- For each pair of morphisms f:A→B and g:B→C, (i.e. cod(f)=dom(g)), a *composite morphism*, g o f: A→C exists (Figure 7.1. C).

The graphical representation of a category can be formalized using the notion of a diagram.

Figure 1. Categorical concepts representation.

The concept of ontology is based on the categorization of things in the real world. Category theory with its logical and analytical features has the potential to be considered as a vehicle for representation of ontologies. An ontology can be viewed in an interconnected hierarchy of theories as a sub-category of a category of theories expressed in a formal logic [14].

Classes in ontology can be defined as a set of properties (attributes and methods) shared by a set of individuals within an equivalence class. Whitmire [12] was one of the few who identified a model based on category theories for object oriented software measurement. Here we follow his approach for demonstration of ontological elements. We define category Class with attribute domains as objects and set-theoretic functions as arrows. In category theory, the cross product of two objects is an object. We can also define some operations for a class. The state St2 is determined by behavior of the object and the state St1 according to the object model for the class of the object model the behavior of an object as a set of operations defined for the class. 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 = <St1, St2> [13] where St1 is the start state and St2 is the end state. Category *Class* is defined with 3 types of objects and 3 types of arrows. The 3 types of objects are [12]:

1- The state space for the class, labeled with the name of the class.

2- The domain sets for the attributes in the class, labeled with the name of the domain.
3- The steady states (a situation in which the relevant variables are constant over time) for objects of the class, labeled with the name for the state used in the application domain.
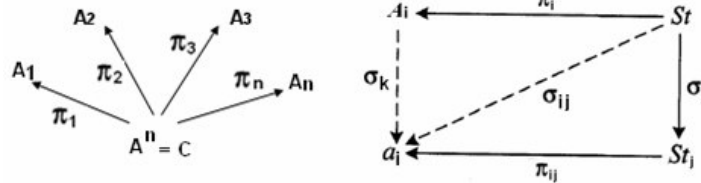The three types of arrows are: projection, selection and an operation arrow for each event.

Figure 2. Representation of the n attribute domains, and the state space of class C (adapted from 12).

The projection arrow for each attribute is drawn from the state space to the attribute domain and labeled with the name of the attribute. The value of the ith attribute is provided by π_i. A selection arrow for each state is drawn from the state space to the state and labeled as σ_x where x is the name of the state. An operation arrow for each event $E = <St1, St2>$ drawn from $S1$ to $S2$ and labeled with the name of the method to which the operation corresponds [12]. One can select a state using the selection function σ_i which gives the ith state.

5. Managing ontological changes using category theory

The categorical representation enables the incremental analysis of ontologies. After describing the ontological concepts within categories representing a modular hierarchy of domain knowledge, we employ category theory to analyze changes in the following ways:

I. By comparing a previous state of a class with a later state: A categorical model [12] is able to describe the 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. It also allows the definition of message passing and method binding mechanisms. Category theory has a special type of mapping between categories called *functor*. Functors are defined as morphisms in the category of all small categories (where classes are defined as categories) [15]. As mentioned above, in our approach, we represent conceptualization of things indexed by times, for example from the FungalWeb Ontology: "*enzyme* has_pH _ optimum at *t*" is rendered as "*enzyme*-at-*t* has_ pH_optimum". Then we use a set of categories indexed by time using functors to capture different state of ontological structure at different time points. The category O at time t that is represented as Ot models the state of the ontologies and all the related interactions at this time. Using a functor allows us to represent the transition from Ot to Ot' (Figure 3) where the time changes from t to t'. In addition, each sub ontology A can be modeled by the series of its successive states At from its '*Creation*' to '*Destruction*' [13].
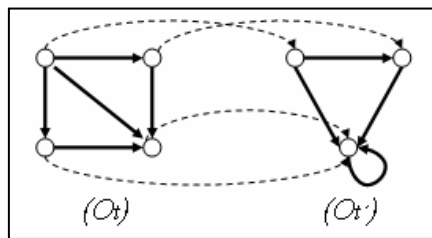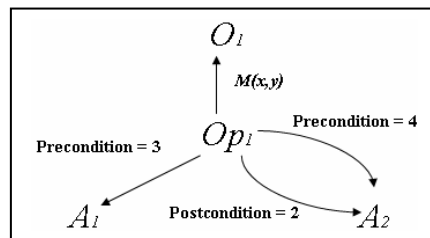

Figure 3. Using Functor.

Figure 4. Measuring Coupling.

II. By measuring coupling: Coupling specifies the extent of the connections between elements of a system and it can identify the complexity of an evolving structure. Measuring coupling is useful for predicting and controlling the scope of changes to an ontological application. Often a change in one class can cause some changes to the dependent classes. When the coupling is high, it indicates existence of a large number of dependencies in an ontological structure which must be checked to analyze and control the chain of changes. Coupling for ontological elements can be described by a number of connections and links between them. So, we focus on arrows in category theory to study these connections. For analyzing a conditional change we followed the formal model described in [12] by identifying three types of arrows in our category: precondition, post-condition and message-send arrows for an existing category [12]. The type of message is determined by the types of changes caused by a method. In the category shown in Figure 4, the coupling for the operation Op_1 is a nonnegative number which can be calculated by the count of the three types of arrows (post-conditions, preconditions and M(x,y)).

6. Application scenario

We are currently applying the proposed methods for managing changes in FungalWeb Ontology and its associated applications. The FungalWeb Ontology is the result of integrating numerous biological database schemas, web accessible textual resources and interviews with domain experts and reusing some existing bio-ontologies. The Ontology is designed with a high level of granularity and implemented in OWL-DL language to take advantage of the combination of a frame representation of OWL and expressiveness of Description Logics (DL). The FungalWeb Ontology links fungal species with enzymes, enzyme substrates, enzyme classifications, enzyme modifications, enzyme related intellectual property, enzyme retail and applications. Figure 5 represents a portion of the FungalWeb application.
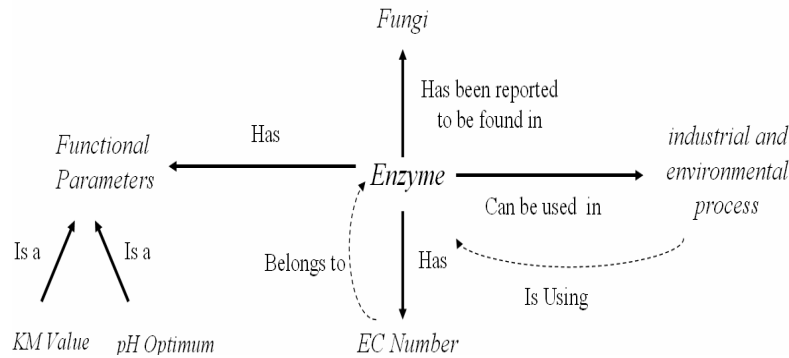

Figure 5. A portion of the FungalWeb Ontology.

Based on our application we designed our class diagrams following the method described in [12] (Figure 6). The Op_i arrows in this figure represent the operations for the class. In this class, the operation or event Op_1 causes an object in state St_1 to transition to state St_2. The operation $Op1$ has no effect upon the object if it is in any other state, since there is no arrow labeled Op_1 which originates in any other state. The object $\varnothing$ in the diagram is the null state. The create arrow represents the creation of the object by assigning an identifier to the object and setting its state to the initial defined state, and the destroy arrow represents its destruction.
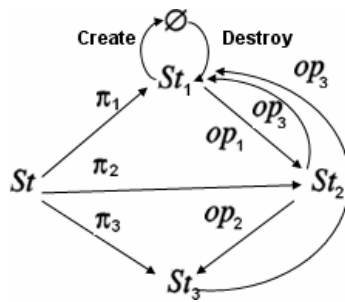
Figure 6. A Class diagram for part of a class structure.

7. Summary and Future Work

The vast amounts of fungal species are still unspecified due to the lack of knowledge. As the knowledge about fungi species grows and new methods become available we can anticipate a fundamental change in the current fungal taxonomy structure. Also the existing fungal classes are promoted, moved, folded, deleted, merged and renamed as more is known about life. We believe category theory has a significant potential to be considered as a supplementary tool to capture and represent the full semantics of ontology driven applications and it can provide a formal basis for analyzing complex evolving biomedical ontologies. For the future research we plan to generalize our usage of category theory along with named graphs and description logics in order to improve ontological conceptualization change management.

8. References

[1] Christopher J.O. Baker, Arash Shaban-Nejad, Xiao Su, Volker Haarslev, Greg Butler, "Semantic Web Infrastructure for Fungal Enzyme Biotechnologists", J. of Web Semantics. Elsevier, 4(3) 2006, 168-180.
[2] Heywood, V.H., ed. "Global Biodiversity Assessment", Cambridge University Press, 1995.
[3] P. W. Crous, "Plant pathology is lost without taxonomy" Outlooks on Pest Management, vol. 16, 2005, 119-123.
[4] Crous PW, Groenewald JZ, Gams W, "Eyespot of cereals revisited: ITS phylogeny reveals new species relationships", European J. Plant Pathol. 2003, 109: 841–50.
[5] Mortensen Chris and Edward N. Zalta (ed.), "Change", The Stanford Encyclopedia of Philosophy, First published Wed Dec 18, 2002; substantive revision Tue Dec 19, 2006.
[6] Hawksworth D.L., "Fungal diversity and its implications for genetic resource collections" Studies in Mycology, 2004, 50: 9–17.
[7] Cassandras, Christos G.; Lafortune, Stéphane, "Introduction to Discrete Event Systems", Kluwer Academic Publishers, September 1999. ISBN 0-7923-8609-4.
[8] A. Schüßler, D. Schwarzott, C. Walker, "A new fungal phylum, the Glomeromycota: phylogeny and evolution", *Mycol. Res*. 105 (12): 1413-1421 (Dec. 2001).
[9] Swann, E.C., E.M. Frieders, and D. J. McLaughlin. 1999. Microbotryum, Kriegeria, and the changing paradigm in basidiomycete classification. *Mycologia* 91: 51-66.
[10] Eilenberg, S. and S. Mac Lane: 1945, "General Theory of Natural Equivalences", Trans. of the American Mathematical Society 58, 231–294.
[11] S. Mac Lane, "Categories for the Working Mathematician", 1971 (corrected 1994), Springer, 1994.
[12] Scott A. Whitmire, "Object Oriented Design Measurement", John Wiley & Sons, September 1997, ISBN: 0-471-13417-1.
[13] A. EC. Ehresmann and J. P. Vanbremeersch, "The Memory Evolutive Systems as a Model of Rosen's Organism-(Metabolic, Replication) Systems", Springer 2006, 16:137–154.
[14] M. J. Healy, T. P. Caudell, "Ontologies and Worlds in Category Theory: Implications for Neural Systems", Axiomathes Journal, 2006, Springer 16:165–214.
[15] Steve Awodey, "Category Theory", Oxford University Press, 2006, ISBN-10: 0-19-856861-4.