# Categorical Representation of Evolving Structure of an Ontology for Clinical Fungus

Arash Shaban-Nejad and Volker Haarslev

Department of Computer Science and Software Engineering, Concordia University,
H3G1M8 Montreal, Quebec, Canada
{arash_sh,haarslev}@cs.concordia.ca

**Abstract.** With increasing popularity of using ontologies, many industrial and clinical applications have employed ontologies as their conceptual backbone. Ontologies try to capture knowledge from a domain of interest and when the knowledge changes, the definitions will be altered. We study change management in the FungalWeb Ontology, which is the result of integrating numerous biological databases and web accessible textual resources. The fungal taxonomy is currently unstable and evolves over time. This evolution can be seen in both nomenclature and the taxonomic structure. In an experiment we have focused on changes in medical species of fungus which can potentially alter the related disease name and description in an integrated clinical system. In order to address certain aspects of representation of changes in an ontology driven clinical application we propose a methodology based on category theory as a mathematical notation, which is independent of a specific choice of ontology language and any particular implementation.

**Keywords:** Bio-Ontologies, Category Theory, Change Management, Fungal Genomics.

## 1 Introduction

Ontologies provide an underlying discipline of modeling medical applications by defining concepts, properties and axioms. They are useful in current medical applications for: sharing common vocabularies, describing semantics of programming interfaces, providing a structure to organize knowledge, reducing development effort for generic tools and systems, improving the data and the tool integration, reusing organizational knowledge [2] and capturing behavioral knowledge. We have implemented the FungalWeb Ontology [1] which is a formal bio-ontology in the domain of in the domain of fungal enzymology with a large number of instances implemented in OWL-DL. We are now trying to develop a change management mechanism 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. Modifying and adjusting ontologies in response to changing data or requirements is not a trivial task. One of the most fundamental questions in our research is: how to represent changes? In order to address certain aspects of representation of changes in an ontology driven

R. Bellazzi, A. Abu-Hanna, and J. Hunter (Eds.): AIME 2007, LNAI 4594, pp. 277–286, 2007.

application in the biomedical domain, in this paper we propose a method based on category theory. In our research, we have focused on ontologies not in isolation but as artifacts that are part of an integrated healthcare system. As an experiment we have focused on changes in medical species of fungus which can potentially alter the related disease name and description in an integrated clinical system.

## 2 Fungi Phylogeny and Evolution

Fungi are widely used in industrial, medical, food and biotechnological applications. They are also related to many human, animal and plant diseases, food spoilage and toxigenesis [4]. Fungi are also interesting because their cells are surprisingly similar to human cells [5]. The reason is that fungi split from animals about 1.538 billion years ago - 9 million years after plants did – therefore fungi are more closely related to animals than to plants [6]. It is estimated that there are about 1.5 million fungal species [7] on the earth, but only about 10% of those are known and only a few of the known fungus have an identified usage such as yeast for making bread, beer, wine, cheese and a few antibiotics [5]. A small percentage of discovered fungi have been linked to human diseases, including dangerous infections. Treating these diseases can be risky because as mentioned above human and fungal cells are very similar. Any medicine that kills the fungus can also damage the human cells. Thus knowing more about fungi and correct identification of each fungi species is crucial and can improve the quality of fungal-based products and also helps to identify new and better ways to treat serious fungal infections in humans. Fungus are also the main source of agricultural and plant diseases, so identifying them will help for tracking and controlling these diseases [5]. Typically, fungal evolution studies have been based on comparative morphology, cell wall composition [8], ultrastructure [9], cellular metabolism [10], and the fossil records [11]. Recently, by advances in cladistic and molecular approaches new insight is provided [12]. Some other new identification methods are based on Immuno-taxonomy and polysaccharides [12], which are highly suited antigens for the identification of fungi at the genus and species levels [13]. The following fungal chemical substances are also used as complementary characters to the classical morphological taxonomy of fungi: proteins, DNA, antigens, carbohydrates, fatty acids and secondary metabolites. One can find a review of the methods for employing the substances in [14]. These substances are very valuable at many taxonomic levels and they play an increasing role in the clarification of the phylogeny (a classification or relationship based on the closeness of evolutionary descent) of fungi [13].

### 2.1 Name Changes in Fungal Taxonomy

Most fungal names are not stable and change with time. Fungal names reflect the data about organisms and as our understanding of the relationships among taxa increases, names will be forced to change so that they do not implicitly contradict the data [15]. Most names are currently based on the phenotype (visible characteristics of organism). As more data become available, however, we run into various problematic issues, such as convergent evolution, seen as the evolution of the same form in different families and even orders, so that similar anamorphs (the imperfect (asexual) state of a

fungus)) may have completely different, unrelated teleomorphs (the sexual stage in the life cycle of a fungus; considered the perfect stage). These names then have to change, as they no longer convey the correct information to the user [15]. 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 [16]. The morphological conceptualization is not sufficient, and will no longer 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 [15]. Another issue in fungal taxonomies is dual nomenclature (two names for one organism) due to the anamorph/teleomorph debate [15]. This is caused by the fact that it is frequently impossible to say when an asexual state belongs to a specific sexual state without the backup of molecular data. A study on revision of the fungi names [17] shows that between 1960 and 1975, 212 names of foliicolous lichenized fungi were described or used by A.C. Batista and co-workers.

### 2.2 Managing Name Changes

We are currently in the middle of a revolution in fungal taxonomy [15]. Names are linked to data. Older names, are mostly classified based on small data sets (mostly phenotypic), and therefore they are subject to change. How biologists can deal with this process of continuous change? To answer to this question 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. By advances in the technology in the future, biologists hope to preserve the fungal taxonomy from change by using unique DNA signatures and species identifier numbers to recognize the species rather than using their name [19]. Currently only around 16% of 100000 known fungal species are represented by DNA sequence data [15], which is approximately 1.1% of the estimated 1.5 million species on Earth, thus it seems that a very low percentage of the already discovered fungal species are in fact being preserved from the change [20]. The changing nomenclature of fungi medical importance is often very confusing. Currently some of the pathogenic fungi have a very unstable taxonomy. For instance, the name of the fungi, Allescheria boydii which can cause various infections in humans, was changed to Petriellidium boydii and then to Pseudallescheria boydii within a short time [23]. Consequently, the infections caused by this organism were referred to as allescheriasis, allescheriosis, petriellidosis, and pseudallescheriosis in the medical literature [24]. In order to manage the changes in fungal names and clarify the ambiguities, the Nomenclature Sub-Committee of the International Society for Human and Animal Mycology (ISHAM) published its regulations for mycosis nomenclature [23, 24]. Based on these regulations a disease should be named, with a meaningful name describing the disease, while in the traditional disease taxonomies the names “fungus+sis” indicate only a causative fungal genus which could be highly influenced by the taxonomic changes. In addition, in the new regulation the value of names of the “pathology A due to fungus B” construction was emphasized [23], e.g., “subcutaneous infection due to *Alternaria longipes*” [12].

### 2.3 Changes and Revisions in Taxonomic Structure

By advancing in molecular biology and changing the fungal nomenclature, one can expect changes in taxonomical structure and relationships. Here are some examples:

**Example 1:** *Glomeromycota* was discovered in 2001 [25] as a new fungal phylum. The arbuscular mycorrhizal (AM) fungi and the endocytobiotic fungus, *Geosiphon pyriformis*, are analyzed phylogenetically by their small subunit rRNA gene sequences. By studying their molecular, morphological and ecological characteristics, it is discovered that they can be separated from all other major fungal groups in a monophyletic clade [25]. Consequently they are removed from the polyphyletic *Zygomycota*, and located into a new monophyletic phylum, the *Glomeromycota* with four new orders *Archaeosporales*, *Paraglomerales*, *Diversisporales* and *Glomerales* [25].

**Example 2:** The sedge parasite *Kriegeria eriophori* has never been satisfactorily classified, because a number of its characters at the gross micromorphological and ultrastructural levels appeared to be autapomorphic [26]. Recently by using the nucleotide sequence data approach which provides more information than standard morphological approaches, some of the ultrastructural characters were discovered to be synapomorphies for a group containing *K. eriophori* and *Microbotryum violaceum.* These characters serve to define the new subclass Microbotryomycetidae [26].

## 3 Category Theory and Ontologies

Category theory is a new domain of mathematics, introduced and formulated in 1945 [27]. A formal model of objects based on "category theory" is introduced in [28]. Employing formalisms based on logics and mathematics in order to move the Web from being only human understandable, to being both human and machine understandable is the known goal of Semantic Web defined by W3C [30]. Category theory is closely connected with computation and logic [31] which allows an ontology engineer to implement different states of design models to represent the reality. Using categories one can recognize certain regularities to distinguish a variety of objects, capture and compose their interactions and differentiate equivalent interactions, identify patterns of interacting objects and extract some invariants in their action, or decompose a complex object in basic components [32]. 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 class of morphisms ("arrows") and for each morphism $f$ there exists one object such as A as the domain of $f$ and one object such as B 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 $f: A \rightarrow B$ and $g: B \rightarrow C$, (i.e. $cod(f) = dom(g)$), a *composite morphism*, $g \circ f$: $A \rightarrow C$ exists (Figure 7.1 (c)).

Representation of a category can be formalized using the notion of a diagram.

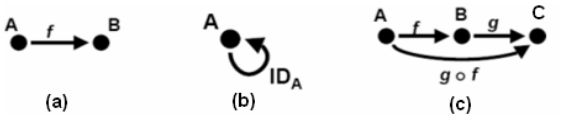


**Fig. 1.** Categorical concepts representation

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 [29]. In fact we can use category theory to represent ontologies as a modular hierarchy of domain knowledge. Ontological relationships represented using category theories are considered to be directed [18] to show the direction of information. These "relationships" are known as "morphism".

### 3.1 The Category Class

Classes can be defined as a set of properties (attributes and methods) shared by a set of individuals within an equivalence class. Whitmire [31] was one of the few who identified a model based on category theories for object oriented applications measurement. Here we follow his approach for demonstration of ontological elements. We can 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. 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 $E = <St1, St2>$ [32]. St1 is the start state and St2 is the end state. St1 and St2 are not necessarily distinct and they might refer to the same state [22] (when an even does not change the state). Category *Class* is defined with 3 types of objects and 3 types of arrows. The 3 types of objects are [31]:

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 domain.

Three types of arrows are: projection, selection and operation arrows.

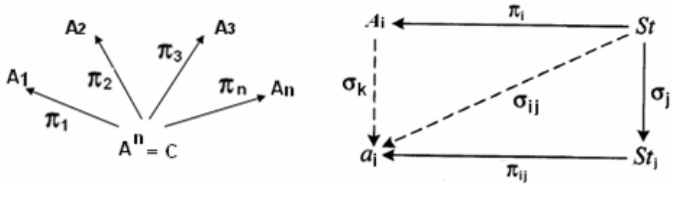


**Fig. 2.** Representation of the n attribute domains, and the state space of class C (adapted from [31])

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 *i*th attribute is provided by $\pi_i$. A selection arrow for each state is drawn from the state space to the state and labeled as $\sigma_x$ 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 [31]. One can select a state using the selection function $\sigma_i$ which gives the *i*th state.

### 3.2 Operations on a Class

Most common operations during ontology evolution are: add a class, delete a class, combine two classes into one, add a generalization relationship, add an association relationship, add/delete a property and add/delete a relationship. Figure 3 represents adding a class to our available structure. Figure 4 (a) and (b) demonstrate adding and dropping a relationship respectively.

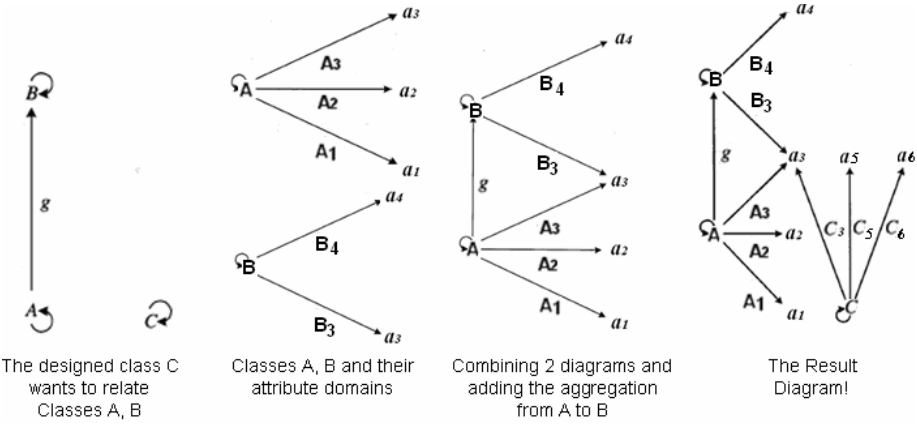


**Fig. 3.** Adding a class to the available structure, based on categorical operation (adapted from [31])

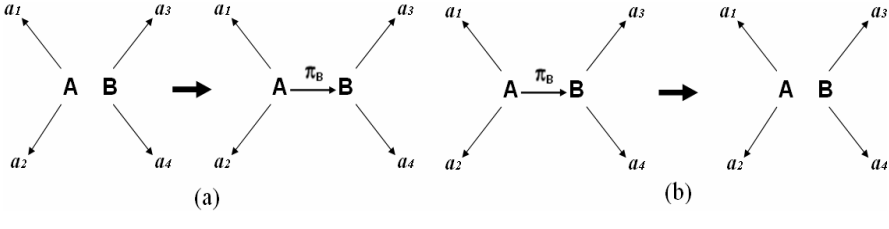


**Fig. 4. (a)** ADD an Aggregation Relationship (**b**) Drop a Relationship [31]

## 4 Managing Changes Using Category Theory

The categorical representation enables the progressive analysis of ontologies. After describing the ontological concepts within categories representing a modular

hierarchy of domain knowledge, we employ category theory to analyze ontological changes in the following ways:

**I. By comparing a previous state of a class with a later state:** A categorical model [31] 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) [21]. The role of time is not usually taken into account in current ontology evolution studies. Considering time in ontologies can increase the complexity and needs a very expressive ontology language to represent it. 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 $O_t$ models the state of the ontologies and all the related interactions at this time. Using a functor allows us to represent the transition from $O_t$ to $O_{t'}$ (Figure 5) where the time changes from *t* to *t'*. In addition, each sub ontology *A* can be modeled by the series of its successive states $A_t$ from its '*Creation*' to '*Destruction*' [32].

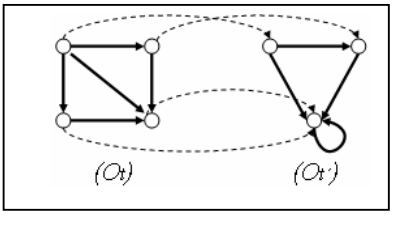


**Fig. 5.** Using Functor

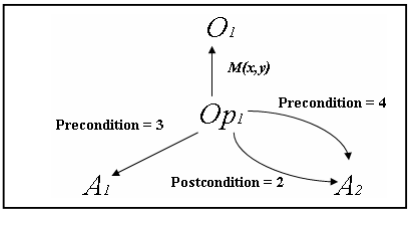


**Fig. 6.** Measuring Coupling

**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 [31] by identifying three types of arrows in our category: precondition, post-condition and message-send arrows for an existing category [31]. The type of message is determined by the types of changes caused by a method. In the category shown in

Figure 6, the coupling for the operation *Op1* is a nonnegative number which can be calculated by the count of the three types of arrows (post-conditions, preconditions and M(x,y)).

## 5 Application Scenario

Bioinformatics is a challenging domain in knowledge management. Biological data are highly dynamic and bioinformatics applications are large and have complex interrelationships between their elements. In addition, they usually have various levels of interpretations for one particular concept. In 1958 Rosen [3] proposed to use category theory in biology, in the frame of a ''relational biology''. At this time, we are applying the proposed methods for managing changes in the FungalWeb Ontology which is the result of integrating numerous biological databases, web accessible textual resources and interviews with domain experts and reusing some existing bio-ontologies. Figure 7 demonstrates a portion of the FungalWeb application in categorical representation.

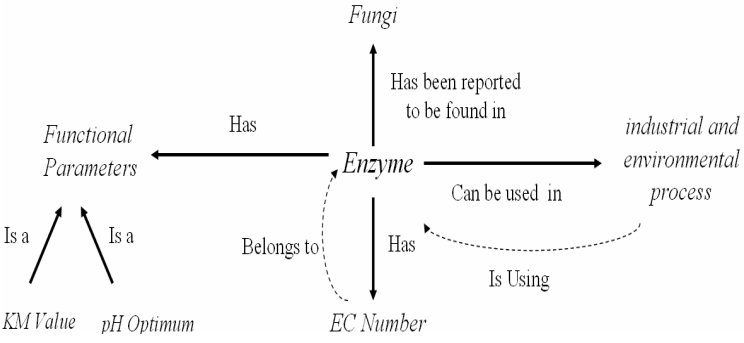


**Fig. 7.** A portion of the FungalWeb application

Based on our application we designed our class diagrams following the method described in [31] (Figure 8). 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 $Op_1$ has no effect upon the object if it is in any other

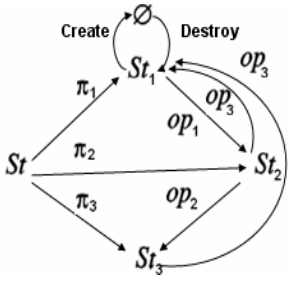


**Fig. 8.** A Class diagram for part of a class structure

state, since there is no arrow labeled Op1 which originates in any other state. The object ∅ 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 destroy arrow represents its destruction.

## 6 Conclusions

As the knowledge about fungi species grows and new methods become available one can anticipate a fundamental change in the current fungal taxonomy structure. 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 other formalisms such as Petri nets, Named graphs and Description Logics in order to improve ontological conceptualization change management. For ontology versioning we also plan to use category theory to determine the degree of semantic similarity between different ontology versions. In addition the work on employing other categorical constructors such as *pushoust* and *pullbacks* for analyzing changes in taxonomical structures is still in progress.

## References

1. Baker, C.J.O., Shaban-Nejad, A., Su, X., Haarslev, V., Butler, G.: Semantic Web Infrastructure for Fungal Enzyme Biotechnologists. Journal of Web Semantics 4(3), 168–180 (2006)
2. Santos, G., Villela, K., Schnaider, L., Rocha, A., Travassos, G.: Building Ontology Based Tools for a Software Development Environment. In: Melnik, G., Holz, H. (eds.) LSO 2004. LNCS, vol. 3096, Springer, Heidelberg (2004)
3. Rosen, R.: The Representation of Biological Systems from the Standpoint of the Theory of Categories. Bulletin of Mathematical Biophysics 20, 245–260 (1958)
4. Bernabé, M., Ahrazem, O., Prieto, A., Leal, J.A.: Evolution of Fungal Polysaccharides F1SS and Proposal of Their Utilisation as Antigenes for Rapid Detection of Fungal Contaminants. E. Journal of Env., Agr. & Food Chem. 1, 30–45 (2002)
5. McLaughlin, D., Rinard, P., Cassutt, M.: Discovery about evolution of fungi has implications for humans. University of Minnesota (20 October, 2006)
6. Nikoh, N., Hayase, N., Iwabe, N., Kuma, K., Miyata, T.: Phylogenetic relationships of the kingdoms Animalia, Plantae and Fungi, inferred from 23 different protein species. Mol. Biol. Evol. 11, 762–768 (1994)
7. Heywood, V.H. (ed.): Global Biodiversity Assessment. Cambridge University Press, Cambridge (1995)
8. Bartnicki-Garcia, S.: The cell wall in fungal evolution. In: Evolutionary biology of the fungi, pp. 389–403. Cambridge University Press, New York (1987)
9. Heath, I.B.: Nuclear division: a marker for protist phylogeny. Prog. Protis. 1, 115–162 (1986)
10. LéJohn, H.B.: Biochemical parameters of fungal phylogenetics. Evol. Biol. 7, 79–125 (1974)

11. Hawksworth, D.L., Kirk, P.M., Sutton, B.C., Pegler, D.N.: Ainsworth and Bisby's dictionary of the fungi, 8th edn. Intern. Myco. Institute, Egham, United Kingdom (1995)
12. Guarro, J., Gene, J., Stchigel, A.M.: Developments in fungal taxonomy. Clinical Microbiology Reviews 12(3), 454–500 (1999)
13. Notermans, S., Dufrenne, J., Wijnands, L.M., Engel, H.: H.J. Med.Vet. Mycol.26, 41–48 (1988)
14. Frisvad, J.C., Bridge, P.D., Arora, D.K.: Fungal chemical taxonomy. Marcel Dekker, Inc., New York-Basel-Hong Kong (1998)
15. Crous, P.W.: Plant pathology is lost without taxonomy. Outlooks on Pest Management 16, 119–123 (2005)
16. Crous, P.W., Groenewald, J.Z., Gams, W.: Eyespot of cereals revisited: ITS phylogeny reveals new species relationships. European J. Plant Pathol. 109, 841–850 (2003)
17. Lucking, R., Serusiaux, E., Maia, L.C., Pereira, E.C.G.: A Revision of the Names of Foliicolous Lichenized Fungi Published by Batista and Co-workers Between 1960 and 1975. The Lichenologist 30(2), 121–191(71) (1998)
18. Krötzsch, M., Hitzler, P., Ehrig, M., Sure, Y.: Category Theory in Ontology Research: Concrete Gain from an Abstract Approach. Technical Report, AIFB, U of Karlsruhe (March 2005)
19. Crous, P.W., Groenewald, J.Z.: Hosts, species and genotypes: opinions versus data. Australasian Plant Pathology 34(4), 463–470 (2005)
20. Hawksworth, D.L.: Fungal diversity and its implications for genetic resource collections. Studies in Mycology 50, 9–17 (2004)
21. Awodey, S.: Category Theory. Oxford University Press, Oxford (2006)
22. Wand, Y.A.: A Proposal for a Formal Model of Objects. In: Kim, W., Lochovsky, F. (eds.) Object-Oriented Concepts, Databases, and Applications, pp. 537–559. ACM Press, New York (1989)
23. Odds, F.C., Arai, T., Di Salvo, A.F., Evans, E.G.V., Hay, R.J., Randhawa, H.S., Rinaldi, M.G., Walsh, T.J.: Nomenclature of fungal diseases, A report from a Sub-Committee of the Intl' Society for Human and Animal Mycology (ISHAM) (1992)
24. Odds, F.C., Rinaldi, M.G.: Nomenclature of fungal diseases. Curr. Top. Med. Mycol. 6, 33–46 (1995)
25. Schüßler, A., Schwarzott, D., Walker, C.: A new fungal phylum, the Glomeromycota: phylog eny and evolution. Mycol. Res. 105(12), 1413–1421 (2001)
26. Swann, E.C., Frieders, E.M., McLaughlin, D.J.: Microbotryum, Kriegeria, and the changing paradigm in basidiomycete classification. Mycologia 91, 51–66 (1999)
27. Eilenberg, S., Mac Lane, S.: General Theory of Natural Equivalences. Transac tions of the American Mathematical Society 58, 231–294 (1945)
28. Mac Lane, S.: Categories for the Working Mathematician (corrected 1994). Springer, Heidelberg (1971)
29. Healy, M.J., Caudell, T.P.: Ontologies and Worlds in Category Theory: Implications for Neural Systems. Axiomathes Journal 16, 165–214 (2006)
30. Caldwell, B., Chisholm, W., Vanderheiden, G., White, J.: Web Content Accessibility Guidelines 2.0. W3C Working Draft 11 March 2004 (2004)
31. Whitmire, S.A.: Object Oriented Design Measurement. John Wiley & Sons, Chichester (1997)
32. Ehresmann, A.EC., Vanbremeersch, J.P.: The Memory Evolutive Systems as a Model of Rosen's Organism-(Metabolic, Replication) Systems, vol. 16, pp. 137–154. Springer, Heidelberg (2006)