Incremental Biomedical Ontology Change Management through Learning Agents

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. Biomedical knowledge bases and ontologies constantly evolve to update the knowledge in the domain of interest. One problem in current change management methodologies is the over-reliance on human factors. Despite the advantages of human intervention in the process of ontology maintenance, including a relative increase of the overall rationality of the system, it does not guarantee reproducible results of a change. To overcome this issue, we propose using intelligent agents to discover and learn patterns for different changes and their consequences. In this paper, we present a novel multi-agent-based approach, to manage the evolving structure of biomedical ontologies. This framework aims to assist and guide ontology engineers through the change management process in general, and aids in tracking and representing the changes, particularly through the use of category theory. It provides an efficient way to automatically capture, validate, and implement a change.

1 Introduction

Bioinformatics is a challenging domain in knowledge management. Biological data are highly dynamic, and the large biomedical knowledge sources contain complexly interrelated elements, with various levels of interpretation. With advances in life science, many features and functionalities must be added to or removed from existing knowledge bases in the biomedical domain. Bio- ontologies and controlled vocabularies evolve continuously to provide meaningful and valid information. Due to complexity and heterogeneity of their structure, changing one component can have wide-ranging, unpredictable effects. Different versions of an ontology may respond differently to queries, so a system based on frequently changing ontologies may yield responses of questionable quality. The automatic representation, validation and reproduction of changes with minimal human intervention raise challenging issues in ontology evolution. Many bio-ontologies are known to be seriously defective, from both terminological and ontological perspectives. Our approach is based on the RLR framework for employing agents in an integrated, ontology-driven infrastructure to capture the pattern of changes and validate the results. To represent changes in ontol-

ogy-driven biomedical applications and formalize interactions between agents, we also propose the use of category theory as a mathematical notation, independent of any specific ontology language or implementation. We have chosen category theory as the main formalism in our framework, because it proved itself as an efficient vehicle to examine the process of structural change in living and evolving systems [30]. As an experiment, we have employed the RLR framework to manage the evolving structure of the Skin disease Ontology (SKDON). Due to ambiguities in medical mycology from the lack of a standard and consistent vocabularies with comprehensible semantic, there is a need [1] for a standard set of vocabularies to support dermatological practice and enhance the accuracy of clinical knowledge management. The SKin-Disease ONtology (SKDON) is an integrated formal ontology based on OWL-DL, and aims to provide a shared and common understanding of the concepts in the domain of medical mycology for dermatologists and other researchers working on life science. In our work, we have concentrated on disorders of the skin and related tissues, such as hair and nail due to fungi. We have designed the SKDON as an essential part of the FungalWeb semantic web infrastructure [2] to gather, retrieve and manage consistent laboratorial and clinical information for patient care. SKDON is created from several distributed resources, including structured/unstructured texts,　online databases, and existing controlled vocabularies, such as MeSH [3], ICD-9 (http://www.cdc.gov/nchs/icd9.htm), SNOMED (http://www.snomed.org/) and Disease database (http://www.diseasesdatabase.com/). Cross referencing between the FungalWeb ontology, SKDON and MeSH "Chemicals & Drugs" category provides valuable information about the disease, the involved fungus and the drugs prescribed. Change in any of the resources can alter the definitions in the target ontology.

2 The RLR Framework

The RLR framework aims to Represent, Legitimate, and Reproduce the changes and their effects (Figure 1). It helps to capture, track, represent, and manage the changes in a formal and consistent way, enabling the system to generate reproducible results.
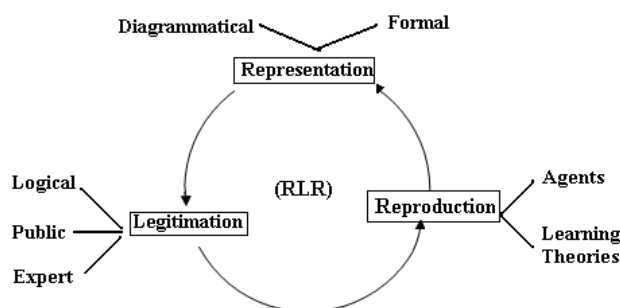

Fig. 1. The RLR framework

In this framework, various ontological changes can be represented in either formal or diagrammatical ways. Each change will be legitimated and validated logically, then

approved publicly and by experts. To reproduce the results of changes and automate the change management process, agents are recruited to learn change patterns and their consequences.

2.1 Representation

This phase is responsible for consistently updating the representations of new knowledge. Many of the problems in ontology evolution are basically problems about the nature and representation of change. For the formal representation of changes, we use Description Logics, and for diagrammatical representation, we employ a method based on discrete state model and category theory [4].

2.2 Legitimation

Legitimation, in our context, is defined as the verification of the legitimacy and consistency of a change in the domain of interest. This phase assesses the impact of a potential change before the change is actually made. Experts and logical reasoners should study a change based on its consistency with the whole design, in various degrees of granularity. Then, the final approval is needed from end-users. Logical legitimation can be obtained by a reasoning agent.

2.3 Reproduction

Over-reliance on human factors is a problem in current change management methodologies. Despite the advantages of in maintenance, including higher rationality, human intervention does not guarantee the reproducibility of results of a change [29]. To overcome this issue, we propose using intelligent agents that discover patterns for different changes and their consequences.
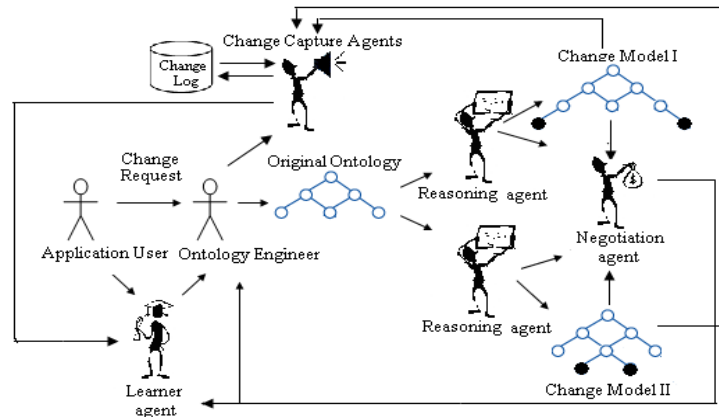

Fig. 2. The change management process using agents.

3 Agents and Pattern of Change

Intelligent agents have the ability to find, identify, and collect desired information from multiple resources about various actions under changing conditions [5]. Agents are able to work rationally in order to capture changes in dynamic and heterogeneous environments, and to respond properly to these changes [6]. In the RLR framework, we have used four types of agents: Change capture agents, Learner agents, Reasoning agents, and Negotiation agents. Figure 2 demonstrates the interactions between them.

3.1 Change Capture Agents

This agent family is responsible for discovering, capturing, and tracking the changes in ontology, by processing one or more change logs. They detect real-world alterations and report them as new facts with which to update the knowledge base of an agent. Changes can occur on a random or scheduled basis. The change capture agents act like triggers in a database. We have defined the following three different types of change-capture agents:

- **Action Control Agents (ACA):** The action control agents consist of user activities and legal operations, which together capture changes such as deletion, insertion, and updates to ontology elements, and can store all the data related to different types of change in change logs.

- **Explorer Agents (EA):** The explorer agents capture changes by processing and reading change logs in parallel, in a specified time range. By logically determining transactions, the explorer agents generate the appropriate messages for the corresponding services.

- **Log-Reading Agents (LRA):** The log-reading agents read the log files in a specified time period. This information will be passed on to a learning agent in order to create patterns for different changes. Later, the information can be used to Undo or Redo a change.

Together, these agents determine which ontological elements have changed. To capture ontological changes, we also use annotation properties such as: Timestamps, Version and Status on ontological elements.

3.2 Learner Agent

As an application is used and evolves over time, the change logs can accumulate invaluable data and information about various types of changes. A learner agent can use these historical records of changes that occur over and over in a change process to derive a pattern. After several changes, possibly from various releases, it would be feasible to estimate the rate and direction of future changes for a system by generating rules or models. In RLR, the reasoner and negotiation agents can change the generated rules, and send modifications to the learning agent. The learning agent starts with

limited, uncertain knowledge of the domain, and tries to improve itself, relying on adaptive learning based on semantics provided by the ontological backbone. The adaptive learner agent plays an important role in the reproduction phase, where we look for patterns to bootstrap the process of change management. The discovery of temporal patterns for event-based data is addressed by P.S. Kam, et al. [7], while Höppner tackled the problem with the discovery of informative temporal rules for defining temporal patterns in [8]. Learning rules for discovering temporal patterns is described by L. Sacchi, et al. [9, 10]. The RLR-Learning agent uses the Sacchi's algorithm [9] for extracting temporal rules to learn patterns of evolving ontological data [9].

3.3 Reasoning Agent

A reasoning agent is a software agent that controls and verifies the logical validity of a system, revealing inconsistencies, hidden dependencies, redundancies, and misclassifications. It automatically notifies users or other agents when new information about the system becomes available. We use RACER [11] as a description logic reasoner agent, along with other semi-formal reasoners in the RLR framework.

When the agent is faced with a change, it ought to revise its conceptualization [12] based on the new input by reasoning about the consistency of the change using both prior and new knowledge. Several attempts [13, 31] have been made, to provide reasoning services for category-based systems. We also use a semi-automated reasoning system for basic category-theoretic reasoning based on a first-order sequent calculus [14]. It captures the basic categorical constructors, functors, and natural transformations, and provides services to check consistency, semantic coherency, and inferencing [14].

3.4 Negotiation Agent

Negotiation happens when agents with conflicting interests desire to cooperate [15]. In the RLR framework, the negotiation agent acts as a mediator allowing the ontology engineer and other autonomous agents to negotiate the proper implementation of a specific change while maximizing the benefits and minimizing the loss caused by such change. A human expert may then browse the results, propose actions and decide whether to confirm, delete, or modify the proposals, in accordance with the intention of the application.

In our framework, negotiation is defined based on the conceptual model of argumentation [16]. In this context, an argument is described as a piece of information that allows an agent to support and justify its negotiation stance or influence that of another agent [15, 17]. Employing argumentation to analyze belief revision with the intention of updating an agent's knowledge also has been studied by M. Capobianco [24], based on dialectical databases.

4 Category Theory for Representing Agent Interactions

Category theory is a new domain of mathematics, introduced in 1945 [18]. Using categories, one can recognize certain regularities in distinguishing a variety of objects, capture and compose their interactions, differentiate equivalent interactions, identify patterns of interacting objects and extract some invariants in their action, and decompose a complex object into its basic components [19]. Categorical notations consist of diagrams with arrows. Each arrow $f: X \rightarrow Y$ represents a function. Representation of a category can be formalized using the notion of diagram.

As presented in [4], category theory has great potential as a mathematical medium to represent, track, and analyze changes in biomedical ontologies. In addition we demonstrated [4] its capability to analyze some of the common operations during ontology evolution. After describing the ontological concepts within categories representing a modular hierarchy of domain knowledge, we have employed category theory as a formalism to analyze ontological changes and agent interaction in different stages of the RLR framework.

Agents perform actions in a context by using rules. The change of the rules is a main adaptation principle [20] for learning in our framework. For describing our adaptive agents we follow G. Resconi's formalization method [20]. Each rule includes a finite or infinite semantic unity symbolized as S_1, IN, P_1 and OUT, respectively representing the input statement, the domain of the rule, the rule, and the range of the rule (denoting the value of an agent's action). Generally, when working in a static environment, we deal with only one family of rules for each context, but when the environment is dynamic, it is very likely that these rules change into other rules. Therefore, a single change in an ontological element triggers other changes in rules and contexts. As an agent gradually learns the different rules for various contexts, a communication channel between these rules and between different agents is necessary. Such changes are demonstrated in [20] as follows (Figure 3).
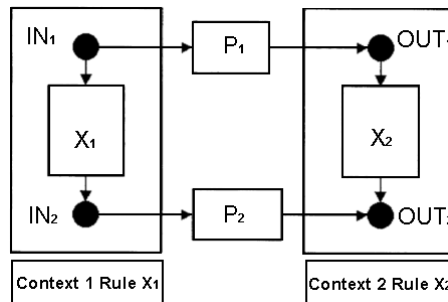

Fig. 3. Demonstration of the semantic unity of the changes of the rule X_1 in the context 1 into the rule X_2 in the context 2 (adapted from [20]).

Agent interactions can be simulated by categories. We have used category theory formalism, along with General Systems Logical Theory (GSLT) explained in [21], to

describe agent communication. For example, the communication between different semantic unities as shown in [20] can be represented as follows:
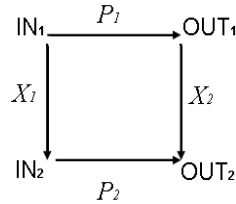
$$IN_1 \xrightarrow{P_1} OUT_1$$

$$X_1 \qquad\qquad X_2$$

$$IN_2 \xrightarrow{P_2} OUT_2$$

Fig. 4. Categorical representation that demonstrates how rules P_1 and P_2 enable the transformation of the rule X_1 into the rule X_2 [20].

Category theory can also be used for modeling agent interaction protocols [22] yielding, a practical image of adaptive learning agents, their semantic unities and adaptation channels [20].

5. Application Scenario

Since skin disorders have been historically categorized by appearance rather than scientific and systematic facts [23], the existing taxonomy must be modified to update the ontological truth. Many terms in current medical mycology vocabularies describing skin disorders originate as verbal descriptions of appearance, foods, people, mythological and religious texts, geographical places, and acronyms [24]. Many names and terms are highly dependent on individual or regional preferences [24], causing redundancy, vagueness, and misclassification in current vocabularies. We study various alterations in fungal taxonomy [4]. An example of the proposed changes are *Trichophyton Soudanense* to *Trichophyton Violaceum*, *Trichophyton megninii* to *Trichophyton rubrum*, and *Trichophyton equinum* to *Trichophyton Tonsurans*.
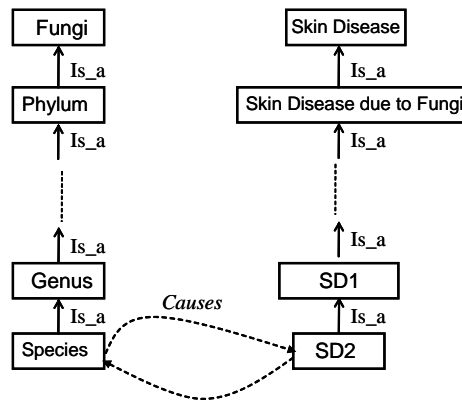
Fig. 5. Changing the fungi name can change the related disease name.

As another example, the pathogenic fungus *Candida glabrata* is now called *Torulopsis glabrata* [26]. Usually changes in fungi taxonomy can alter the related disease name and description (Figure 5). For instance, the name of the fungus, *Allescheria boydii* which can cause various infections in humans, was changed to *Petriellidium boydii* and then to *Pseudallescheria boydii* within a short time [25]. Consequently, the infections caused by this organism were referred to as *allescheriasis*, *allescheriosis*, *petriellidosis*, and *pseudallescheriosis* in the medical literature [28]. In our domain, we need to update and improve the ontological structure of the FungalWeb and SKDON Ontologies regularly for the annotation of fungal genes and analyzing the role of the fungi species in various diseases. For example, the older version of the FungalWeb Ontology did not have sufficient terminology to annotate genes involved in *Malassezia* infections. To meet this new requirement, the updated version of the ontology has gained 26 additional terms addressing these infections.

As it is shown in Section 4, category theory can be used in RLR to represent agent interactions. It is also used as formalism for analyzing changes in FungalWeb Ontology and SKDON. Changes to any part of the ontologies may cause the conceptual design changes the state and also may cause alterations to other dependent artifacts. To represent different states of our conceptualization, we use a categorical discrete state-model, which describes states and events in diagrammatical notion.
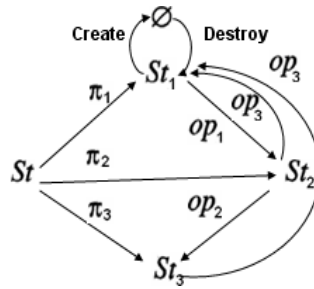

Fig. 6. A Class diagram that represents the transition between states [4].

Based on our application we designed our class diagrams following the method described in [27, 4] (Figure 6), which can be used to create patterns for learning agents. The Op_i arrows in this figure represent the operations for the class, wherein the operation or event Op_1 causes an object in state St_1 to undergo a transition to state St_2. The operation Op_1 has no effect upon the object if it is in any other state, since no arrow labeled Op_1 originates from 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 and setting its state to the initial defined state, and the destroy arrow represents its destruction.

6 Discussion and Challenges

Ontologies in general must change to update their ontological 'truth'. The heterogeneity of biomedical ontologies and the volatility of their knowledge sources increase the odds of different structural alterations. One issue in the domain of ontology evolution is lack of formal change models with clear, comprehensible semantics. In this manuscript we have introduced a novel multi-agent framework to handle changes in bio-ontologies based on category theory. This framework assists an ontology engineer to capture, track, represent and manage the changes in a formal and consistent way which enables the system to create reproducible results. Using category theory with its dynamic nature in our model allows capturing the full semantics of evolving bio-ontologies as well as providing a formal basis to represent agent interactions.

In the process of employing category theory as the core formalism for the RLR framework, we had to deal with a variety of challenges, including the reasoning issues and management of conceptualization changes. However, we are able to provide basic reasoning and inferencing for categories, though we still must improve the reasoning capability to cover more advanced services. The representation of conceptualization changes is another challenge, especially for abstract concepts and notions. To overcome this, we are working on grammatical change algorithms in linguistics and language evolution. Minimizing human intervention is an issue in the "Reproduction" phase, although improvement of the learning and negotiation algorithms for the agents may reduce the problem.

References

1. Ackerman, B.: Need for a Complete Dictionary of Dermatology Early in the 21st Century. Arch Dermatol.136: 23 (2000)
2. 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)
3. National Library of Medicine, Medical Subject Headings, Bethesda, MD, 1994.
4. Shaban-Nejad, A., Haarslev, V.: Categorical Representation of Evolving Structure of an Ontology for Clinical Fungus. In: Bellazzi, R., Abu-Hanna, A., Hunter, J. (eds.) AIME 2007. LNCS, vol. 4594, pp. 277--286. Springer, Heidelberg (2007)
5. Devedžic, V.: Knowledge Modeling - State of the Art. Integrated Computer-Aided Engineering 8(3), 257--281(2001)
6. Li, L., Wu, B., Yang, Y.: Agent-Based Approach for Dynamic Ontology Management. KES (3) 1--7 (2005)
7. Kam, P.S., Fu, A.W.: Discovering temporal patterns for interval-based events. In: Kambayashi, Y., Mohania, M.K.,Tjoa, A.M. (eds.) DaWaK 2000. LNCS, vol. 1874, pp. 317—326. Springer, Heidelberg (2000)
8. Höppner, F.: Knowledge discovery from sequential data. PhD thesis, Technical University Braunschweig, Germany, (2003)
9. Sacchi, L., Bellazzi, R., Larizza, C., Porreca, R., Magni, P.: Learning Rules with Complex Temporal Patterns in Biomedical Domains. In: Miksch, S., Hunter, J., Keravnou, E.T. (eds.) AIME 2005. LNCS, vol. 3581, pp. 23--32. Springer, Heidelberg (2005)
10. Sacchi, L., Larizza, C., Combi, C., Bellazzi, R.: Data mining with Temporal Abstractions: learning rules from time series. Data Mining Knowledge Discovery 15(2) 217--247 (2007)
11. Haarslev, V., Möller, R.: Description Logics for the Semantic Web: Racer as a Basis for Building Agent Systems. KI 17(3) 10--15 (2003)

12. Capobianco, M., Chesñevar, C.I., Simari, G.R.: Argumentation and the Dynamics of War-ranted Beliefs in Changing Environments. Journal of Autonomous Agents and Multi-Agent Systems 11(2) 127--151 (2005)

13. Poigné, A.: Elements of Categorical Reasoning: Products and Coproducts and some other (Co-) Limits. In: Pitt, D.H., Abramsky, S., Poigné, A., Rydeheard, D.E. (eds.) CTCS 1985. LNCS, vol. 240, pp. 16-42. Springer, Heidelberg (1986)

14. Kozen, D., Kreitz, C., Richter, E.: Automating Proofs in Category Theory. In: Furbach, U., Shankar, N. (eds.) IJCAR 2006. LNCS, vol. 4130, pp. 392--407. Springer, Heidelberg (2006)

15. Rahwan, I., Ramchurn, S.D., Jennings, N.R., McBurney, P., Parsons, S., Sonenberg, L.: Argumentation Based Negotiation. Knowledge Engineering Review 18(4) 343--375 (2003)

16. Van Eemeren, F.H., Grootendorst, R.F., Henkemans, F.S.: Fundamentals of Argumentation Theory: A Handbook of Historical Backgrounds and Contemporary Applications, L. Erl-baum Associates, NJ, USA (1996)

17. Jennings, N.R., Parsons, S., Noriega, P., Sierra, C.: On argumentation-based negotiation. In: International Workshop on Multi-Agent Systems (IWMAS98), Boston, USA (1998)

18. Eilenberg, S., Mac Lane, S.: General Theory of Natural Equivalences', Transactions of the American Mathematical Society 58, 231--294 (1945)

19. Ehresmann, A.EC., Vanbremeersch, J.P.: The Memory Evolutive Systems as a Model of Rosen's Organism-(Metabolic, Replication) Systems. Springer, vol. 16, 137--154 (2006)

20. Resconi, G., Jain, L.C.: Intelligent Agents: theory and applications. Vol. 155 of Studies in Fuzziness and Soft Computing, Springer-Ver. Berlin (2004)

21. Resconi, G., Hill, G.: The Language of General Systems Logical Theory: A Categorical View. In: 3rd European Congress on Systems Science, Rome 1091--1096 (1996)

22. Johnson, M.W., Miller, T., McBurney, P.: Modelling Multi-agent Interaction Protocols using Categories Enriched over Pointed Sets. Technical Report, U of Liverpool (2007)

23. Papier, A., Chalmers, R.J.G., Byrnes, J.A. and Goldsmith, L.A.: Framework for improved communication: the Dermatology Lexicon Project. Journal of the American Academy of Dermatology. 50 (4), 630--634 (2004)

24. Al-Aboud, K., Al-Hawsawi, K., Ramesh, V., Al-Aboud, D. and AL-Githami, A.: An Ap-praisal of Terms Used in Dermatology. SKINmed 2 (3), 151--153 (2003)

25. 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)

26. Cushion, M.T., Stringer, J.R.: Has the Name Really Been Changed? It Has for Most Re-searchers. Clinical Infectious Diseases, 41, 1756--1758 (2005)

27. Whitmire, S.A.: Object Oriented Design Measurement, John Wiley & Sons, New York, NY, USA (1997)

28. Odds, F.C., Rinaldi, M.G.: Nomenclature of fungal diseases. Curr. Top. Med. Mycol. 6, 33 -- 46 (1995)

29. Flouris, G.: On Belief Change and Ontology Evolution. Ph.D. thesis in the Department of Computer Science, University of Crete (2006)

30. MacFarlane, A.I.: Dynamic structure theory: A structural approach to social and biological systems. Bulletin of Mathematical Biology. 43(5), 579--591 (1981)

31. Girard, J.Y., Lafont, Y., Taylor, P.: Proofs and Types. Cambridge University Press (1989)