Strategic Health Information Management and Forecast: The BirdWatching Approach

Arash Shaban-Nejad[1] and Volker Haarslev[1]

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

Abstract. To facilitate communication and the exchange of information between patients, nurses, lab technicians, health insurers, physicians, policy makers, and existing knowledge-based systems, a set of shared standard terminologies and controlled vocabularies are necessary. In modern health information management systems, these vocabularies are defined within formal representations called ontologies, where terminologies are only meaningful once linked to a descriptive dataset. When the datasets and their conveyed knowledge are changed, the ontological structure is altered accordingly. Despite the importance of this topic, the problem of managing evolving ontological structures is inadequately addressed by available tools and algorithms, partly because handling ontological change is not a purely computational affair. In this paper, we propose a framework inspired by a social activity, birdwatching. Using this model, the evolving ontological structures can be monitored and analyzed based on their state at a given time. Moreover, patterns of changes can be derived and used to predict and approximate a system's behavior based on potential future changes.

Keywords: Change management, Biomedical ontologies, Multi-agent system, Health information management.

1 Introduction

> *"When you know what the habitat and the habits of birds are watching them is so much more interesting."*
>
> *The Beginners Guide to Bird Watching[1]*

Strategic information systems (SIS) are widely used in the healthcare industry to support real-time decision making and consistent maintenance of various changes in strategic vision. Many strategic information systems have employed various controlled vocabularies, ontologies, and knowledge bases as their conceptual backbone to stan-

---

[1] http://birdwatchingforbeginners.info/

dardize and facilitate human-human, human-agent, and agent-agent interactions and communications (Figure1).

Using biological classification and clinical vocabularies/lexicons has a long history in medicine and life science dating back to Aristotle's *scala naturae* [1] (scale of nature), which was a very simple method of dividing organisms into groups, ranging from the simple species to more complex ones, based on their appearance. In the 17th century, Carl Linnaeus, who is often referred as the father of modern taxonomy, developed his classification system for the naming and classification of all organisms. Linnaeus represented his classification method based on binomial nomenclature (e.g., humans are identified by the binomial *Homo sapiens*). Later, as the understanding of the relationships between organisms changed, taxonomists converted the five ranks into the seven-rank hierarchy by adding the two ranks of "Phylum" (between Kingdom and Class) and "Family" (between Order and Genus).
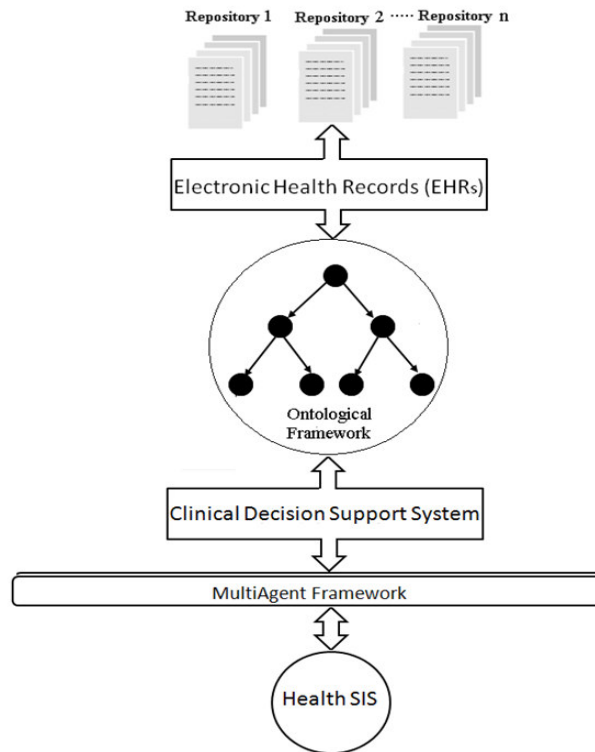

Fig. 1. An abstract representation of the interactions in a typical knowledge-based Health Strategic Information System (SIS).

Change in the taxonomic ranks is still an ongoing process. Due to advances in knowledge and the influence of evolutionary techniques as the mechanism of biological diversity and species formation, taxonomists needed a new classification scheme to

reflect the phylogeny of organisms. Also, recruitment of new criteria, besides structural similarities, such as genetic codes and molecular features, and advances in tools and techniques resulted in the discovery of various organisms, altering the older structures and forming new kingdoms with new branches and terminologies [2]. The biomedical classifications and terminologies have been organized in several models [3] as Controlled Vocabularies, Thesauri, Taxonomies, and Ontologies.

A relatively new trend is emerging to use ontology, as defined by Gruber [4] ("specification of conceptualization"), to provide an underlying discipline of sharing knowledge and modeling biomedical applications by defining concepts, properties and axioms. Modifying and adjusting ontologies in response to changing data or requirements are significant barriers to the implementation of efficient biomedical ontologies in real clinical environments. Depending on the size and complexity of the ontological structures, their maintenance can be very expensive and time consuming. In this paper, we introduce the sociotechnical aspects of our agent-based framework, which aims to assist and guide ontology engineers through the change management process in evolving biomedical ontologies [5] with minimal human intervention.

2   The BirdWatching: A Nature Inspired Approach

Since the existing biomedical knowledge bases are being used in various organizational and geographical levels (i.e. institutional, local, regional, national and international), any change management framework should be able to address this decentralization and distribution nature. One of the critical tasks in any change management framework is *traceability,* which provides transparent access to different versions of an evolving system. It also aids in understanding the impact of a change, recognizing a change and alerting upon occurrence, improving the visibility, reliability, auditability, and verifiability of the system, propagating a change [6], and reproducing results for (or undoing effects of) a particular type of change. Advances in impact analysis gained by traceability facilitate predictability in the post-change analysis stage in an ontology maintenance framework.

To explain our method for change management more intuitively we use a conceptual metaphor based on Birdwatching activity. Birdwatching as a recreational and social activity is the process of observation and study of birds through a particular time frame using different auditory devices. Figure 2 shows a sequence[2] of typical activities recommended for Birdwatching.

---

[2] Bird Watching Tips for Beginners: http://animals.about.com/od/birding/tp/birdidtips.htm
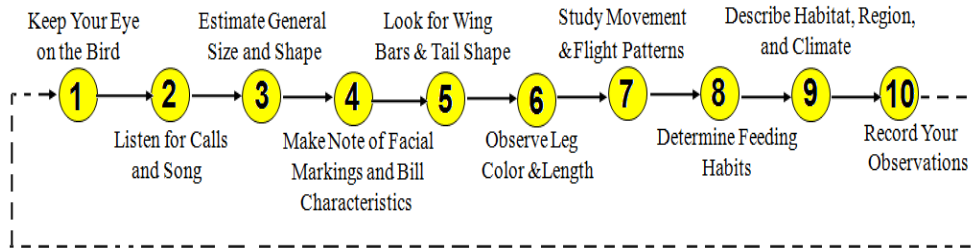
Fig. 2. A series of activities in Birdwatching.

Looking at the list of activities presented in Figure 2 one can discover that the central idea of Birdwatching, which is tracking the position of the birds at different time points and predicting their path by deriving a flight pattern based on recorded observed information, is quite close in spirit to monitoring any dynamic spatial-temporal system. Inspired by this metaphor we have designed a multi-agents framework called RLR [7], which aims for Representation, Legitimation and Reproduction of changes in ontological structures. Using intelligent agents reduces several issues related to human intervention in dynamic e-health systems [8].
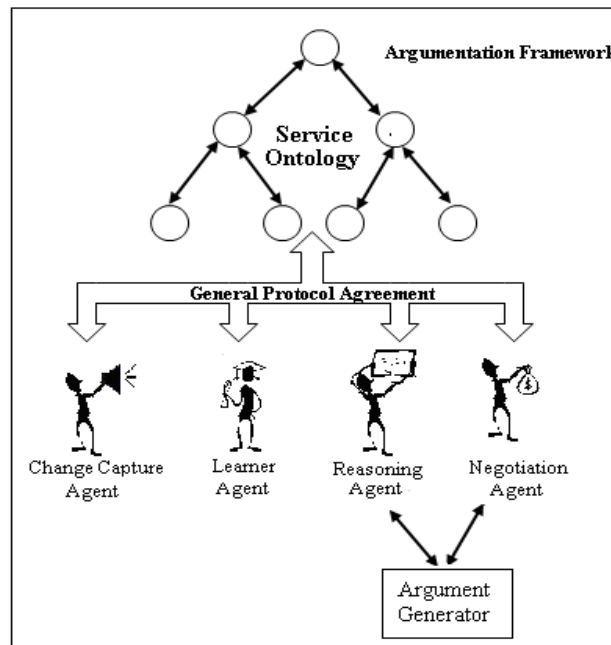

Fig. 3. The RLR framework with a service ontology providing consensus between agents. To reach an agreement among the agents and provide a common understanding, the service ontology is needed, so that updating this ontology generates a new understanding for the software agents, which can then update and adjust their beliefs based on new knowledge.

As part of RLR, we have defined a set of change capture agents, learning agents, negotiation agents, and reasoning agents within an argumentation-based framework (Figure 3) that enables agents with conflicting interests to cooperate. To reach an agreement among the agents and provide a common understanding between them, a service ontology, as shown in Figure 3, is needed so that updating this ontology generates a new understanding for the software agents, which can then update and adjust their beliefs based on new knowledge. Employing service ontologies to automatically provide a service profile to describe the supported services and the related communicative transactions and invoke the services for service-seeking agents is currently being considered as a solution to overcome some of the issues related to overreliance on human intervention. However, these ontologies will not remain static and unchanged throughout their life cycle, and managing their dynamic structure would be part of the whole problem itself.
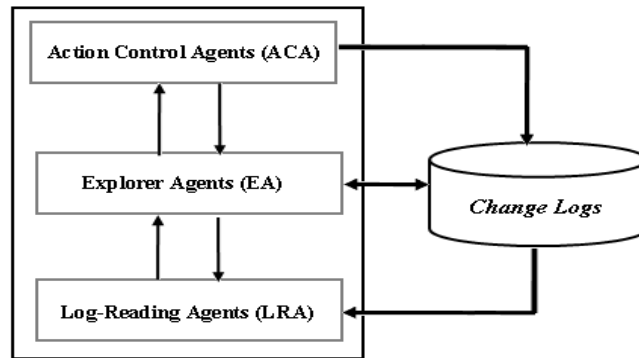

Fig. 4. The cooperation between the change capture agents.

In the RLR framework the change capture Agents (Figure 4) are responsible for Tasks 1 and 2, represented in Figure 2. The changes logs store the information about the changes (Task 4). The learning agents start with limited knowledge (Task 5 and 6) and improve themselves by gaining inferred knowledge (Tasks 8 and 9) based on the semantics provided by the ontological backbone. Moreover, the learning agents along with negotiation agents (which manage the negotiation process to find a proper way to implement a change) and reasoning agents (which check for inconsistencies and perform final validations) can derive a pattern of change using the information stored in the change logs and the background and derived knowledge (Task 7). Using this pattern one can achieve a practical estimate for expected changes (Task 3). Finally the result of the observation will be stored to be used for future inferencings (Task 10), and to choose an appropriate pattern (Task 7) in the reproduction phase.

The final outcome, which has been generated through a rigorous argumentation process over generally accepted arguments, has an implicit link to the archived historical processes that can be reused to choose a proper pattern (Task 7) in the reproduction phase (Figure 5).
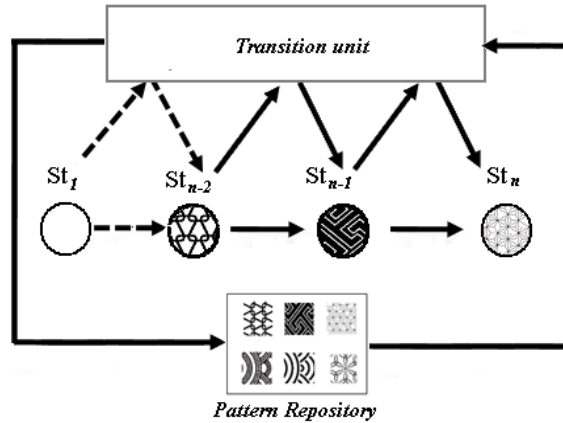
Fig. 5. A generic transition system in a multi-agent system. A system changes its state from St_1 to St_2 via a transition unit and a rigorous argumentation process between the agents to choose proper patterns from the change pattern repository for implementing a certain type of change.

3 Formalization of the Framework

Looking at the different tasks in Birdwatching one can discover that the central idea of Birdwatching is quite close in spirit to monitoring any dynamic spatial-temporal system. The Galileo's dialogue [24] for explaining motion for the first time stated that for capturing and tracking a moving object one needs to record the position of that object in each instance of time.

3.1 Categorical Representation

Here we use category theory [9], as an algebraic notation independent of any implementation language, to study the ontological dynamism by mapping from a category of times to a category of states or back to our Birdwatching metaphor, the bird's flight (motion) can be represented by mapping from a category of times to a category of spaces (Figure 6). 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 and we use categorical constructors for capturing the states of ontologies at different time points.
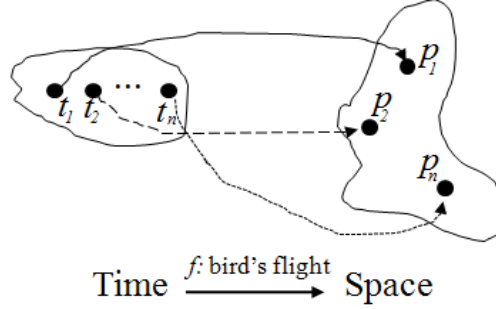
$$\text{Time} \xrightarrow{\ f:\text{ bird's flight}\ } \text{Space}$$

Fig 6. A map from category of time points to category of positions in space for describing a bird's flight categorical perspective [25].

Similarly, the behavior of an individual ontological element (state) can be monitored by function g, which maps the time points to the set of positions for the element in the ontology.

$$\text{Time} \xrightarrow{\ g:element's\_behavior\ } \text{Ontology}$$

Moreover an ontology has different states and behaves in a distributed semantic web environment.

$$\text{State} \xleftarrow{\ i:has\_state\ } \text{Ontology} \xrightarrow{\ h:behavies\ } \text{Semantic Web}$$

Composing these diagrams one can see that a behavior of an individual ontological element should be studied in close relations with time, the state and the behavior of the whole ontological structure in a semantic web environment (Figure 7).
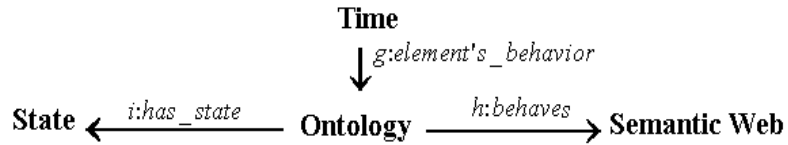

Fig. 7. A temporal diagram for studying the behavior of ontologies.

Using category theory enables us to formally represent and track the evolving ontological structure and the argumentation network. It also provides a formal basis to be used by the RLR system for recommendations and conflict resolution. We have also employed the category theoretical distributed graph transformation techniques [10] to analyze the model transition and transformation using certain conditions, which are specified via transformation rules. As an example, in Figure 8, consider two taxonomies related to ontologies O_1 (source ontology) and O_2 (target ontology), where each node represents a concept, which is identified with a label along with a set of corresponding attributes. After discovering similarities and differences between these two taxonomies, we need to find a proper transformation that has been transformed O_1 to O_2. To start this procedure, the two taxonomies need to be aligned and brought into a

mutual agreement, based on the matching concepts (the ones that affected less in the transformation) within the ontologies. The matching will be computed based on the degree of similarities between two concepts. From the categorical point of view, the problem of comparing two hierarchical structures can be studied by exploring isomorphisms in their structures.
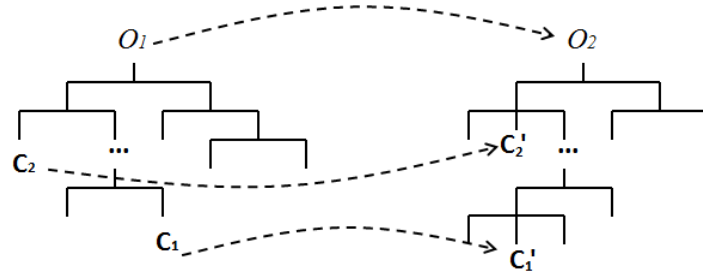

Fig. 8. The alignments between some concepts in two ontologies O_1 and O_2.

The use of graph transformations helps us discover the set of operations that transforms the hierarchy indicating the old version of an ontology into the hierarchy indicating the new one.

3.2 The RLR Dialectic Change Management

In debates on distinguishing between "Dependent" and "Independent" entities in the real world, the two concepts of *Ontological Philosophy* and *Dialectic Change* attracted our attention. The concept of Ontological Philosophy [11] focuses on the wholeness and unity of the world and considers change as an aspect of substances in the real world. From the other side, the concept of Dialectical Change [12] tries to represent a change as new forms built upon the old and by combining the new and the old without total replacement, implying both newness and continuity. In this theory, any change needs a cause and can be placed through a process. Holsti [12] used the Marxist idiom, the synthesis, as a metaphor for this processes.

Using the concept of "*Dialectic Change*" as a metaphor, we can introduce our formal agent-based argumentative framework, where "synthesis" takes place, for studying ontology evolution and shifting as model transformation. This transformation results from quantitative changes accumulated over a period of time and generates a new form out of old patterns ("coexistence of both old and new") [12]. In fact, most of the changes that occur in an ontological structure, which lead to a new state, emerge from the preceding states[3]. In other words, the change lies within the system [13]. Therefore, "learning" about different actions in different states of a system seems to be a key factor for starting a successful change management mechanism.

---

[3] A "state" in this manuscript is being used to express a situation describing a part of the real (dynamic) world in a specific instance of time.

3.3 Models of Learning

By determining the tradeoffs between losses and benefits that can result from agents' actions, we will be able to have a mathematical model to foresee the agents' (software or human) behavior. A state of "Nash equilibrium" [14] is one of the popular approaches in evolutionary game theory for modeling the most beneficial (or least harmful) set of actions for a set of intelligent agents. For the sake of prediction, Nash equilibrium can be understood as "a potential stable point of a dynamic adjustment process in which individuals adjust their behavior to that of the other players in the game, searching for strategy choices that will give them better results" [15]. In our framework the intelligent agents decide on the proper actions and able to change and improve their decisions based on what they learn. Here, following the approach given by [23], for each learning agent, we define an internal state; a function that shows how an agent decides and chooses actions based on its internal state (decision-making); the functions showing the payoff dominance (loss/benefit); and a state update function, specifying how an agent updates its state based on the payoff received from previous iterations. The state of each agent depends on the probability distribution over all the possible situations [23], and the one with the highest probability can specify the final decision. Another technique for automating the learning process is through inductive bias. The inductive bias of learning [16] in neural networks is a set of assumptions, given as input, that the learner uses to predict and approximate the target outputs (even for unseen situations) through a series of training instances and their generalization.

3.4 Anomaly Pattern Analysis

Intelligent agents in RLR also detect and generate patterns of anomalies, either syntactic or semantic, by assessing and analyzing consistent common errors that occur through different revisions. After the anomalies have been flagged by change capture agents, the learner agent can then be taught the proper route for performing the revisions through a set of pattern mining algorithms (see [17] as an example of techniques for mining dynamic patterns). This task is crucial in a wide variety of applications, such as biosurveillance for disease outbreak detection [18] and cancer diagnosis. The learner agents not only enable the RLR framework to manage potential, expected, and prescheduled changes, but also prepare it for dealing with random and unexpected alterations. However, human supervision and participation will be anticipated for the former case.

3.5 The Change Analysis Model in RLR

Our change analysis model is composed of a set of states that are linked to their predecessors and successors through some defined relationships. This allows us to check backward and forward compatibilities for one specific ontological structure from a given state. This is determined by defining various conditions and constraints for an

event. The conditions can later be used to restore the previous state based on the insights gained for each event. Somehow it means a revision or review of the past, or an attempt to define an alternate (parallel) past [19]. Since ontological assertions are based on open world assumptions, neither past nor future knowledge about the world is complete. One can always ask questions (e.g., "Could a specific mutation, under certain circumstances, lead to the species X or Y?") and draw a different path from the previous states to the subsequent states. This iterative process of switching between the future, current, and revised past states has been regarded in [19] as the process of "rolling back to some previous state and then reasoning forward" in the form of queries such as, "Is there some future time in which p is true?" [19].

To deal with forward and backward compatibility, in our research we have employed graph transformation techniques, which enable us to analyze different states of the graphs based on the given initial states and the transformation rules. Indeed, graph transformation offers many benefits, but lacks sufficient expressivity and semantics to deal with all aspects of ontology change management. Our approach for this issue can be improved by recruiting a formal mathematical representation such as category theory. The enhancement can be done in two aspects: 1) the transformation rules can impose restrictions on ontology transformation in the way that, for example, some alteration can be prohibited, or some changes, which have less impact on ontological elements, can be excluded in the related change analysis (e.g., the transition of a fungus from one genus to another does not affect its physical appearance); 2) the changes in states can be scheduled to occur simultaneously, sequentially, or in parallel.

3.6 Identity Preservation in RLR

The identity of a concept can be determined by those properties and facts that remain stable through time, even during multiple ontological changes. If ontologies are able to maintain their conceptual stability, they can better preserve their intended truth. To this end, the RLR framework employs a defensive mechanism to prevent harmful changes and reduce the risk of potentially dangerous actions by incrementally adapting to the changes at different levels. If a destructive change is about to happen in the ontology (e.g., deleting a concept, such as "fungi", when other dependent concepts, such as "fungal infection", exist), a warning signal will be sent to the agents based on the knowledge within the ontology (e.g., "fungi are the cause of fungal infections") to infer the potential threat and prepare them to plan for a proper action. This mechanism works much like the self-awareness system inside rational animals, which helps them avoid possible dangers without actually experiencing their life threatening influences. For example, as pointed out in [20], a person who is confronted with fire does not have to experience the burning sensation and can run away as a counteraction, since the person has been taught that smoke indicates fire and that fire can kill humans.

4 Discussion and Future Works

To avoid the fatal errors caused by uncontrolled changes in biomedical knowledge-based systems, a consistent change management process with minimum human intervention is vital. In this paper, we have described a method based on a metaphor taken from a recreational activity, birdwatching, to highlight the temporal aspects of ontologies by representing the conceptualization of things indexed by times, which enables one to control forward and backward compatibilities for taxonomic revisions. In fact, our introduced approach, based on the insights from category theory, can be employed to develop algorithms and tools to assist ontology change management. In our recent experiments, we have applied the introduced agent-based method, formalized with category theory, in several biomedical applications, including the management of requirement volatility in e-health systems [21] and analyzing the evolutionary relationships between fungal species [22]. Currently, we are working to improve our rule-based graph transformation method and extend it to cover hierarchical distributed graphs, which support nested hierarchies in different levels of abstraction.

References

1. Verhagen, F.C.: Worldviews and Metaphors in the human-nature relationships: An Ecolinguistic Exploration Through the Ages. Language & Ecology, vol. 2 no. 3 (2008)
2. eHealth: standardized terminology. Executive Board 118th Session, EB118, Provisional agenda 8.4., 25 May, 2006, World Health Organization.
   http://apps.who.int/gb/ebwha/pdf_files/EB118/B118_8-en.pdf
3. Hedden, H.: Controlled Vocabularies, Thesauri, and Taxonomies. The Indexer, 26(1): 33--34 (2008)
4. Gruber, T.R.: A translation approach to portable ontologies. Knowledge Acquisition, 5(2): 199--220 (1993)
5. Shaban-Nejad, A., Haarslev, V.: Bio-medical Ontologies Maintenance and Change Management. In: Sidhu, A.S., Dillon, T.S. (eds.) Biomedical Data and Applications. Studies in Computational Intelligence, vol. 224, pp. 143--168, Springer (2009)
6. Smith, M.J., Dewar, R.G., Kowalczykiewicz, K., Weiss, D.: Towards Automated Change Propagation; the value of traceability. Technical Report, Heriot Watt University (2003)
7. Shaban-Nejad, A., Haarslev, V.: Incremental biomedical ontology change management through learning agents. In: Nguyen, N.T., Jo, G., Howlett, R.J., Jain, L.C. (eds.) KES-AMSTA 2008. LNCS, vol. 4953, pp. 526--535. Springer (2008)
8. Shaban-Nejad, A., Haarslev, V.: Human Factors in Dynamic E-Health Systems and Digital Libraries. Pease, W., Cooper, M., Gururajan, R. (eds.) Biomedical Knowledge Management: Infrastructures and Processes for E-Health Systems. Information Science Reference – ISR, pp. 192--203. IGI Global (2010)
9. Asperti, A., Longo, G.: Categories, types, and structures: an introduction to category theory for the working computer scientist. MIT Press, Cambridge, MA (1991)
10. Ehrig, H., Ehrig, K., Prange, U., Taentzer, G.: Fundamentals of Algebraic Graph Transformation. Monographs in Theoretical Computer Science. An EATCS Series, Springer, New York (2006)
11. Scribner, P.: Introduction to Ontological Philosophy (1999),
    http://www.twow.net/MclOtaI.htm

12. Holsti, K.J.: The Problem of Change in International Relations Theory. Paper No. 26 from CIR Working Paper Series, (1998)
13. Gilbert, M.C.: The Dialectics of Knowledge Management (2006) http://news.gilbert.org/DialecticsKM
14. Osborne, M.J.: Nash Equilibrium: Theory. In: An Introduction to Game Theory, Oxford University Press, USA (2003)
15. Holt, C.A., Roth, A.E.: The Nash equilibrium: A perspective. PNAS, 101(12): 3999--4002 (2004)
16. Mitchell, T.M.: The need for biases in learning generalizations. In: Shavlik, J.W., Dietterich, T.G. (eds.) Readings in machine learning. pp. 184--191. Morgan Kaufmann (1990).
17. Chung, S., McLeod, D.: Dynamic Pattern Mining: An Incremental Data Clustering Approach. J. Data Semantics, 2: 85--112 (2005)
18. Wong, W.K., Moore, A.W., Cooper, C.F., Wagner, M.: Bayesian Network Anomaly Pattern Detection for Disease Outbreaks. In: 20th International Conference on Machine Learning (ICML2003), pp. 808--815, AAAI Press (2003)
19. Mays, E.: A Modal Temporal Logic for Reasoning about Change. In: 21st Annual Meeting of the Association for Computational Linguistics (ACL), pp. 38--43, Cambridge, MA, US, (1983)
20. Heylighen, F.: Representation and Change. A Metarepresentational Framework for the Foundations of Physical and Cognitive Science. Communication & Cognition, Gent, 200 p. (1990)
21. Shaban-Nejad, A., Ormandjieva, O., Kassab, M., Haarslev, V.: Managing Requirements Volatility in an Ontology-Driven Clinical LIMS Using Category Theory. International Journal of Telemedicine and Applications, Article ID 917826, 14 p. (2009) doi: 10.1155/2009/917826
22. Shaban-Nejad, A., Haarslev, V.: Ontology-inferred phylogeny reconstruction for analyzing the evolutionary relationships between species: Ontological inference versus cladistics. In: 8th IEEE International Conference on Bioinformatics and Bioengineering (BIBE 2008), Athens, Greece, pp. 1--7, IEEE Press (2008)
23. Wang, J.: Computational Approaches to Linguistic consensus. Dissertation at University of Illinois at Urbana-Champaign. (2006)
24. Galileo, G.: Dialogue Concerning the Two Chief Systems of the World - Ptolemaic and Copernican. (1632) http://www.gap-system.org/~history/Extras/Galileo_Dialogue.html
25. Lawvere, F.W., Schanuel, S.H.: Conceptual Mathematics: A First Introduction to Categories. 2nd edition, Cambridge University Press. (2009) (The 1st ed. published on 1997)