THE LIVE ONTOLOGY MODEL

AN AUTONOMOUS AND ADAPTIVE ARCHITECTURE FOR KNOWLEDGE EMERGENCE

Eldon Marks

Department of Computer Science
Faculty of Natural Sciences
University of Guyana

eldon.marks@uog.edu.gy

Abstract

The ontology is considered to be a technological step towards someday realizing the “Semantic Web” in which machines are able to manage the sharing, combination and retrieval of knowledge across the information space of the World Wide Web. The dynamic and emergent nature of knowledge, however, is one of the most challenging problems faced in current Semantic Web researches. The representational primitives and constrained definitions of the ontology make for an architecturally limited, inflexible and somewhat inappropriate structure when considering the nature of knowledge. The distillation of background researches reveals three predominant properties of interest for the architectural advancement of the ontology, namely, autonomy, adaptiveness and emergence. Autonomy suggests the marginalization of human dependency during the ontology lifecycle. Adaptiveness suggests the capability to accommodate new or evolving knowledge during operation. Emergence suggests that knowledge artifacts should arise from real-time distillation of information rather than constrained definitions and specifications. This paper documents the Live Ontology Model: a multicellularity-inspired agent-based model for autonomous and adaptive knowledge emergence.

Keywords: ontology, agent-based model, multicellularity, agent, knowledge modelling, semantic web, knowledge emergence

1. Introduction

The nature of today's knowledge-economy demands that machines accumulate, organize and render vast amounts of knowledge on our behalf. According to the DIKW (Data, Information, Knowledge, Wisdom) hierarchy [17], [1], [4], knowledge is metaphorically viewed as the next, distilled "link in the chain" following information and trailing wisdom (see Figure 1). In this hierarchy, information is data with meaning whereas knowledge is derived from an organised, inter-related collection of information [6].

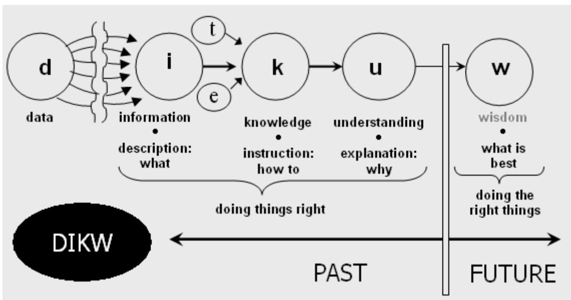

Figure 1 - The DIKW Chain

Continuously advancing hardware and software capacity has provided an effective means to process, generate, store and share information to the point at which the vastness of our growing electronic information space makes human-level processing impractical when further distilling information into readily accessible knowledge. In this regard, researchers endeavour to extend the current hardware and software capacity of machines to autonomously organise information into knowledge. Aligned with this effort, the ontology emerged from the field of Artificial Intelligence as a machine-compatible means to model a knowledge domain. The use of representational primitives and constrained definitions of an ontology, however, make for an architecturally limited, inflexible and somewhat inappropriate structure when considering the dynamic and amorphous nature of knowledge. This structural mismatch becomes evident when applying the traditional ontology to knowledge representation, particularly in scientific and industrial domains. With ongoing discoveries, revisions and retractions, ontology maintenance is absolutely necessary, however it is regarded as a daunting task, often involving a large collection of ontology engineers, dedicated experts and many man-hours [9]. This paper proposes a multicellularity-inspired approach to model knowledge at the machine level in a manner that is architecturally adaptable, autonomous, emergent and consequently better aligned with knowledge modeling requirements than the traditional ontology approach.

1.1 Towards Knowledge Modelling

The demand for machine-level modeling of "distilled information" has prompted a worldwide collaborative movement led by the W3C (World Wide Web Consortium). The W3C's yet-to-be-realised "Semantic Web" is one in which machines are able to manage the sharing, combination and retrieval of information on a semantic level within the massive, heterogeneous information space of the Web. Various languages have emerged from this effort including the Resource Description Framework (RDF), Web Ontology

Language (OWL), SPARQL Protocol and RDF Query Language (SPARQL) and XML. These languages are used to describe ontologies or query the resulting knowledge representations on machines.

An ontology is formally described as a specification of a conceptualization [5], consisting of a representational vocabulary used to constrain definitions of classes, relationships and functions based on a specific domain. While the Semantic Web vision is yet to be largely realised, various consumers have employed ontologies directly as needed. The ontology has become increasingly popular as a tool to structurally organise information and relate large amounts of data in a logical fashion [15]. Ontologies have demonstrated particular usefulness in creating large taxonomies to classify a variety of items such as websites on Yahoo! or products on Amazon. Further, in the Medical field, they have been used as a standard way to share and annotate information among domain experts as in the case of the Unified Medical Language System.

1.2 Ontologies and the Nature of Knowledge

Ontologies however, are not without their shortcomings, which stem from their apparent inability to cope with the complex, emergent and evolving nature of knowledge. The dynamic nature of knowledge is one of the most challenging problems faced in current Semantic Web researches [9]. McInerney [7] alludes to the dynamic characteristic of knowledge in the following statement:

> *"knowledge exists in sentient beings, or emanates from them, and thus it is always changing with the human experience".*

Knowledge has been identified to be a significantly different 'entity' to either information or data [6]. Whereas information and data are entirely finite, measurable and tangible, knowledge is both ephemeral as well as tangible, both tacit and explicit. Tacit knowledge is subjective, tied to context, process-oriented and quite difficult to communicate [6], [10]. Explicit knowledge, on the other hand may be expressed, encoded and transferred. It is explicit knowledge that most current knowledge modeling approaches try to capture, store and render. McInereney, however, suggests that the expression of knowledge is the combination of the behaviour, as well as the artifact, both process as well as object, respectively represented in the tacit and explicit aspects of knowledge. The dynamic nature of knowledge is arguably a result of the subjective activity or processes within the tacit aspect of knowledge, which, in turn renders the explicit artifacts of knowledge.

The process responsible for the rendered artifacts of knowledge allude to yet another quality of knowledge - emergence. Emergence is the property found in resultant complex systems, which are the product of a multitude of simple interactions at fundamental levels. For example, biology can be viewed as an emergent property of the laws of chemistry, which, in turn can be viewed as an emergent property of the laws of particle physics. Likewise and commensurate with the DIKW hierarchy, knowledge can be viewed as the emergent property of information and information, in turn, can be viewed as the emergent property of data [6].

The nature of knowledge, however, as described in the preceding distinguishes it from information and data as an entity that is active by virtue of its emergent and dynamic qualities. Consequently, knowledge requires a means of representation commensurate with its nature when considering machine-level knowledge modelling. McInerney highlights that traditional machine-based structures such as databases may be appropriate to hold information but the nature of knowledge precludes compatibility with these structures. The design premise for typical machine-based structures such as databases is based on the organisation of data from a structural standpoint through the use of defined specifications. Defined specifications, however, are difficult to evolve and maintain beyond their initial design parameters making them unsuitable for representing actively evolving and emergent content such as knowledge.

Interestingly, the design premise of the ontology has been described as an extension of logical database design [13]. While the ontology possesses a richer information model than database schema, over sixty percent of the features are common to both [16]. In general, both are built on strict specifications that cater to the structural representation of content. These specifications form the basic building blocks of ontologies and database schemas.

McInerney's position on the unsuitability of typical machine-based structures for the modelling of knowledge becomes credible in light of the overheads encountered when applying the ontology approach in knowledge representation. Scientific and industrial domains in particular are subject to change as discoveries, revisions and retractions are continuously made in domain knowledge [9]. For the specification of an ontology to remain current in light of the dynamic nature of domain knowledge, periodic maintenance is required which encompasses the extraction of irrelevant information, the introduction of new information and the reinstallation of ontological vocabularies. This process of ontology maintenance may not be feasible in some scenarios and becomes a daunting task in others, often involving a large collection of ontology engineers and dedicated experts.

1.3 Ontology Advancements

Researches in the area of ontology advancement have primarily focused on attempts to automate ontology maintenance as well as related processes such as ontology mergers and alignments, knowledge extraction and knowledge integration. Current solutions are, at best, either semi-automatic with the requirement of a human ontology expert or are only applicable in the initial stages of ontology maintenance.

The work by Srinivasan and Huang [14], presented a radical shift in ontology advancement. Instead of focusing on the refinement of ontological specifications or the development of adjunct, automated processes, they attempted to "liquefy" the ontological structure to create a more adaptive formation. Srinivasan and Huang explored the concept of "fluid ontologies" for organising and browsing knowledge in a digital museum. The basis of their fluid ontology design was to permit knowledge structures to emerge from interactions with content creators of their digital museum. Other key elements of their design are as follows: (1) metaview sharing, whereby the browsing and rearranging of knowledge by participants are explicitly shared; (2) adaptiveness, which permits the ontology to adapt and be redesigned in-time; (3) bots and personalisation, to track and analyse user interaction histories to further fine-tune the dynamic evolution of the ontology. The resulting implementation provided personalised views of knowledge and bot-driven recommendations of the repository for an intuitive and dynamic arrangement of cultural and artistic museum knowledge. While the fluid ontology implementation may have been largely human-dependent for its operation and applied to a digital museum, the authors presented a powerful approach to removing the rigidity of the typical ontology. One of the most notable characteristics of their model is that the fluid ontology is a collection of processes and tools, rather than specifications that permit the emergence of knowledge from data and information.

2. The Live Ontology Model

The distillation of background researches reveals three predominant properties of interest for ontology advancement, namely, **autonomy, adaptiveness** and **emergence**. Autonomy suggests the marginalisation of human dependency during the ontology lifecycle. Adaptiveness suggests the capability to accommodate new or evolving knowledge during operation. Finally, emergence suggests that knowledge artifacts should arise from relatively simple processes involving the distillation of information and data rather than predefined specifications and classifications.

In the following, we take the aforementioned properties into consideration and present the founding insight and design of our Live Ontology Model as an autonomous and adaptive architecture for knowledge emergence.

2.1 Multicellularity: A Design Insight

It is observed that multicellularity demonstrates a tremendous capacity for autonomy, adaptiveness and emergence. At the highest level, multicellular organisms such as animals are viewed to be a complex collection of systems, which are fundamentally composed of simple, specialised cells. Within the body of the organism, there are many organ systems, each of which comprise several organs. Each organ, in turn, is made up of several tissues and each tissue consists of numerous cells; all of which are similar in structure and function. The modular and intricate arrangements of cells, tissues, organs and organ systems permit the organism to be highly adaptive in structure and function; further, synergistic collaborations throughout the structural levels of the organism are made possible through autonomous behaviour originating from the cellular level.

Animal cells come in various shapes, sizes and specilaisations. Their variety and specialisations render a structurally intricate organism with similarly intricate processes and behaviour. The single cell, however, is relatively simplistic both in structure and behaviour. Each specilaised cell has a finite set of behaviours, namely, divide, grow, interact and die. Once a cell has come into being mainly through division, it must grow and reach a state of maturity before it is able to support metabolic processes. Cells must communicate with other cells, respond to stimuli as well as regulate activities to maintain a balance or a state of homeostasis. Finally, once a cell has aged or becomes damaged, it dies naturally and another cell of its kind takes its place.

2.1.1 Multicellular Emergence

The structural complexities of multicellular organisms begin to emerge when groups of cells are formed. Specialized cells arrange themselves to form major classes of tissues, which in turn constitute major organs. An organ may be considered as a component that serves a particular function within the organism. For instance, the lung is designed specifically to transfer gasses between the atmosphere and the blood stream of an organism. At the level of organs, the structural collection and interactions of specialised cells begins to shape the major systems of the organism.

2.1.2 Multicellular Autonomy

The collection of cells in various tissues of an organism is not merely an aggregation of cells but behaves as a community governed by the concept of division of labour. A number of cells participate in a single activity, for instance, muscle cells enable the heart to pump blood and nerve cells transmit nerve impulses. Biologists describe the synergistic collaboration of cells as being likened onto a baton race in which individual runners of a team pass the baton in sequence. Further a cell may divide, grow, interact or die while its adjacent cells hold different physiological states. In this regard the cell operates with a level of autonomy, both as an individual and as part of a community.

2.1.3 Multicellular Adaptiveness

The aforementioned levels of autonomy in cells combined with their modular arrangements permit multicellular organisms to operate with a level of redundancy and flexibility. Worn-out, damaged or unused cells may be replaced or altered while the organism continues to function. In this regard, multicellular organisms possess a higher potential for adaptability than unicellular organisms [12]. Further, through cell specialisation, multicellular organisms are able to maintain a state of homeostasis in spite of changing environmental states. At the cellular level, individual cells are viewed as self-supporting chambers, which act to maintain a steady internal chemical composition in order for metabolic reactions to occur. At higher levels, communities of specialised cells such as epithelial cells form an active layer, which supports homeostasis within the internal environment of the organism in relation to its changing external environment. This cellular formation and behaviour have permitted organisms to develop defenses and adapt to changes both internally and externally throughout the organism's lifetime and lineage.

2.2 The Live Ontology Model Design

The design focus of our Live Ontology Model is a strategic abstraction of multicellualrity at both the structural and behavioural levels. The aim is to effectively emulate, in a knowledge modelling approach, the mechanisms responsible for the autonomous, adaptive and emergent characteristics found in multicellularity. Our method of abstracting and emulating multicellularity is further shaped by the documented characteristics of knowledge as well as relationships among data, information and knowledge of the DIKW hierarchy for a solution that autonomously and adaptively manages data, information and emergent knowledge artifacts.

2.2.1 The Foundation of the Model

The Live Ontology Model is founded on a computational model. Considering the emergent and modular properties of multicellularity, the agent-based computational model (ABM) [8], [3] was found to be a structural and behavioural ideal for our solution.

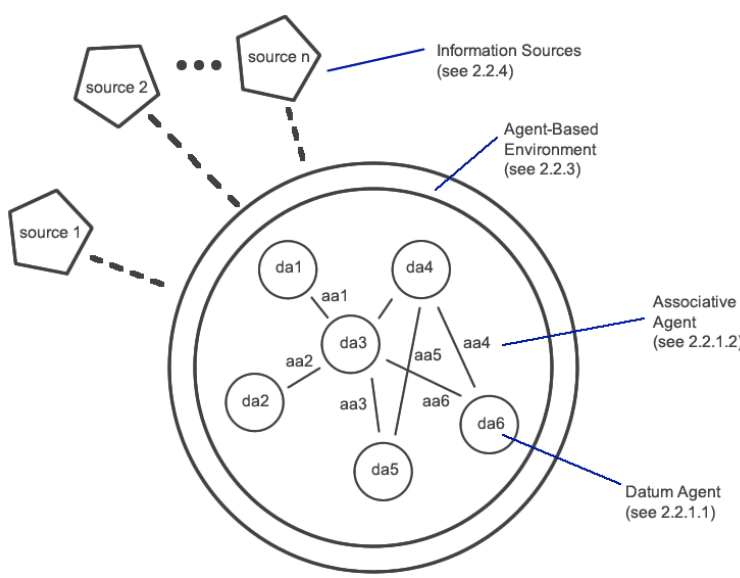

Figure 2 - The Live Ontology Model

The agent-based model is a rule-based computational model in which collections of computational objects or agents interact purposefully, choosing their actions, strategies and locations with some objective in mind. Agent-based models are particularly used in Biological applications to generate populations of entities as well as simulate their interactions in a virtual world. The agent-based model was applied to the Live Ontology Model design particularly due to the following ABM properties:

(1) A modular structure: The resulting behavior of the entire agent-based model is determined by the rule sets of each agent. In addition, individual agent rules sets may be modified and agents may be added or removed without disrupting the entire model [11]. In this regard, it is observed that the modular property of the agent-based model is analogous in structure and behaviour to the identified design of cells in multicellularity. Individual rule sets of modellled agent entities permit autonomous operation while the afforded modularity permits adaptiveness.

(2) Emergence: The localised behaviour of agents governed by their individual rule sets results in more intricate behaviour or emergent patterns as a whole [11]. The multicellular organism as a whole can be described as possessing intricate structure and behaviour made possible by the combinations of relatively simplistic cellular interactions. Further, knowledge can also be described as the intricate outcome that ultimately originates from the collection of relatively simplistic combinations of data according to the DIKW hierarchy.

(3) Stochasticity: In the event of apparent, unpredictable or random behaviour, a probability measure of the given behaviour may be calculated and translated into rules for individual agents [2]. The stochastic property in the agent-based model permits the modelling of non-deterministic and dynamic behaviour as encountered not only in biological systems but in the knowledge modelling as well.

(4) Abstraction: Agent-based models may be constructed with partial or abstracted details of the system under study to produce a simple, verifiable model [11]. Unlike mathematical models, the agent-based model does not require explicit details for provability or completeness. In this regard, only aspects, which are considered most relevant, are incorporated and eventually translated into program logic for execution. Accommodating abstract representations of multicellularity shaped by the DIKW hierarchy is central to the design of the Live Ontology Model.

2.2.2 Agents of the Model

The autonomous agent is fundamental to the agent-based model. Based on abstractions of multicellularity and the DIKW hierarchy, we derived the following classes of autonomous agents for our Live Ontology Model, namely, (1) the datum agent and (2) the associative agent. We arrived at these fundamental agents by observing the relationships and processes that exist among data, information and knowledge. We reduced these relationships and processes to a recursive occurrence of objects forming selective associations with other objects at various levels of abstraction. At the lowest level of abstraction, data are selectively associated to form information; in turn, information is selectively associated to form knowledge. Fundamental to this formation are the object and the association. It can be argued that neither information nor knowledge are singular objects, but rather the result of a collection of objects held together by associative bonds that take on more intricate arrangements towards the level of knowledge. In this regard, the datum and the association were identified as the most fundamental elements of our model.

2.2.1.1 The Datum Agent

The datum agent is designed to represent a single piece of data within the model. In addition, the datum agent is designed with the abstracted behaviour of a single cell in a multicellular organism to foster the qualities of autonomy, adaptiveness and emergence within the model. The datum agent has the following defined goals:

(1) Achieve an operational identity - Analogous to a cell achieving maturity, the datum agent must acquire an essential perception of its own representation in order to achieve its remaining goals. For example, if the datum agent represents "goldfish" then it must acquire an understanding of its own representation particularly in relation to other possible representations. In order to achieve this, the datum agent must consult with a complete source of information external to itself and the environment of pre-existing agents. Within our model, this source of information may be a corpus, wiki or any other reliable electronic collection of information. The datum agent representing "goldfish" may access from this source of information : "a goldfish is a type of fish which lives in a pond". With this information about itself, the datum agent seeks out and engages other datum agents within the environment and establishes associations with them through association agents.

(2) Seek out connections with and engage other datum agents – Once a datum agent has achieved its operational identity, it continuously establishes associations with other datum agents according to the external source of information available to the environment. For example, the "goldfish" datum agent will now seek out and establish associations with the "fish" and "pond" datum agents if they have not already established a similar association or if there are no more possible associations to be formed according to the external source of information.

(3) Propagate when necessary – In the event that a datum agent, after achieving an operational identity, fails to identify one or more of its associated representations in the form of another datum agent, it will "divide" in essence and spawn other datum and association agents, each with partial representations of its own identity. Once the missing agents are spawned, the original datum agent uses the propagated association agents to establish associations with them.

(4) ***Expire when orphaned*** – As the external source(s) of information evolve or are changed, so do the datum and associative agents. In the event that a datum agent loses all associations with other datum agents through changing source information, then it will cease to exist.

2.2.1.2 The Associative Agent

An associative agent is spawned when a connection is established between two datum agents. In essence, the associative agent is a derivative of the datum agent with a scope of interaction limited to the interconnected pair of datum agents. In particular, the associative agent captures and evaluates the active relationship between one datum agent and another. This relationship is limited to one context and is further described by a logical direction between the pair of datum agents, for instance, left to right, right to left or both. Once initialised with an associative context between two datum agents, the associative agent strives to achieve the following goals:

(1) ***Expand on the initial operational identity*** – The associative agent will strive to expand its initial representation by capturing, in parallel, synonymous representations under a single context and direction. For instance, when an associative agent is spawned to connect "goldfish" and "pond" with "goldfish live in ponds", it continues by accessing the external information source to identify comparable relationships between the two datum agents such as "a pond is the habitat of the goldfish". Through this behaviour, the established relationship will carry several dimensions to support a dynamic association between a pair of datum agents.

(2) ***Monitor and evaluate the active association*** – To actively respond to evolving information, the associative agent periodically evaluates its active association against the external information source and adapts accordingly.

(3) ***Expire when orphaned*** – In the event that the associative agent fails to validate its current relationship with either of its datum agents, it ceases to exist.

2.2.3 The Environment of the Model

The agent-based environment provides the declarative services needed by all agents as well as those needed to interface with the knowledge model. The declarative services provided are:

(1) ***Agent creation*** – The environment creates the initial agent(s) and facilitates subsequent "spawn" requests sent by active agents as needed.

(2) ***Agent referencing*** – The environment maintains a registry of all agents which facilitates goal (2) of the datum agent (*Seek out connections with and engage other datum agents*). The agent referencing service creates a fully observable environment in which agents may be aware of each other for goal-based interactions.

(3) ***Source access*** – All agents must be able to access external sources of information from which the knowledge base is built. The model's environment permits multiple, external text-based information sources to be attached at once and made available to all agents of the environment. In addition, the environment provides a level of abstraction allowing agents to utilise a standard method of access and querying regardless of the nature of the attached information source being accessed.

2.2.4 The Information Source of the Model

The information source represents a homogeneous or heterogeneous collection of fact-based, textual information from which the resultant Live Ontology is built. Due to the level of abstraction provided by the agent-based environment, it is theoretically possible to adapt a variety of text-based information sources to the needs of the model. Structured content, however, would be most suitable since this provides the most amenable form, which marginalises or eliminates requisite preprocessing. One such structured information source, which presents a wealth of information, is DBPedia. DBPedia allows the querying of relationships and properties associated with Wikipedia resources available on the World Wide Web. Further, through the attachable and abstracted information source management mechanism of the model, existing ontologies may be assimilated individually to preserve a homogeneous knowledge base or combined to permit a heterogeneous, domain-independent emergence of knowledge.

3. Conclusions and Future Work

The Live Ontology Model is an initial conception and application of autonomy, adaptiveness and emergence as identified properties for ontology advancement. The agent-based model coupled with abstracted insights from multicellularity offers an architecture, which is theoretically capable of autonomous knowledge integration, knowledge adaptation and domain knowledge mergers through combined information sources. The novel approach of modelling the fundamental objects of knowledge - data, as autonomous agents capable of propagation presents the exciting prospect of knowledge expansion as well. Future efforts in this study involve the implementation and evaluation of the Live Ontology Model.

4. References

[1] Ackoff, R.L. (1989). From Data to Wisdom. Journal of Applied Systems Analysis, Volume 16, 1989 p 3-9.

[2] An G., Mi Q., Dutta-Moscato J., Vodovotz Y. (2009). Agent-based models in translational systems biology. Systems Biology and Medicine, 1(September/October), 159-171.

[3] Bonabeau, E. (2002). Agent-based modeling: methods and techniques for simulating human systems. Proceedings of the National Academy of Sciences, 99 (3), pp. 7280–7287

[4] Clark D. (2004). Understanding and Performance. Retrieved from: http://www.nwlink.com/~donclark/performance/understanding.html.

[5] Gruber, T.R. A translation approach to portable ontologies. Knowledge Acquisition, 5(2): pp. 199-220, 1993.

[6] Hey, J. (2004) The Data, Information, Knowledge, Wisdom Chain: The Metaphorical Link. Technical report, University of Berkeley, 2004.

[7] McInerney, C. (2002). Knowledge Management and the Dynamic Nature of Knowledge. Journal of the American Society for Information Science and Technology, 53(12) pp. 1009-1018.
[8] Niazi, M.; Hussain, A. (2009). Agent based Tools for Modeling and Simulation of Self-Organization in Peer-to-Peer, Ad-Hoc and other Complex Networks. IEEE Communications Magazine 47 (3): 163–173.

[9] Nov´aˇcek, V., Laera, L., Handschuh, S., Davis, B. (2008). Infrastructure for Dynamic Knowledge Integration – Automated Biomedical Ontology Extension Using Textual Resources. J Biomedical Informatics 41(5), 816-828.

[10] Nonaka, I., Takeuchi, H. (1995). The Knowledge-Creating Company. Oxford University Press, New York.

[11] Politopoulos, I. (2007). Review and Analysis of Agent-based Models in Biology. Accessed from http://www.csc.liv.ac.uk/research/techreports/tr2007/ulcs-07-021.pdf

[12] Rastogi, V. B., (1997). Modern Biology. Pitambar Publishing, pp. 148-150.
Sharma, N. (2004). The Origin of the Data, Information, Knowledge, Wisdom Hierarchy. Retrieved from: http://go.webassistant.com/wa/upload/users/u1000057/webpage_10248.html

[13] Siricharoen, W. (2009). Ontology Modeling and Object Modeling in Software Engineering. International Journal of Software Engineering and Its Applications 3(1), pp. 43-60.

[14] Srinivasan, R., Huang, J. (2004). Fluid Ontologies for Digital Museums. Technical Report, Harvard University.

[15] Taylor, M. E., Matuzek, C., Klimt, B., Witbrock, M. (2007). Autonomous Classification of Knowledge into an Ontology. In the 20th International FLAIRS Conference (FLAIRS), Key West, Florida, May 2007.

[16] Uschold, M. (2011). Ontologies and Database Schema: What's the Difference?. Retrieved from: http://www.slideshare.net/UscholdM/ontologies-and-db-schema-whats-the-difference

[17] Zeleny, M. (1987). Management Support Systems: Towards Integrated Knowledge Management. Human Systems Management, 7(1987)1, pp. 59-70.