ONTOLOGIES AND COMMUNITIES CO-EVOLUTION IN INFORMATION SYSTEMS

Francesca Arcelli Fontana, Ferrante Formato and Remo Pareschi
University of Milano Bicocca, Milan, Italy
University of Sannio, Benevento, Italy
University of Molise, Campobasso, Italy

Keywords: Communities, Complex networks, Ontologies, Knowledge management.

Abstract: Communities and ontologies are both concepts that have acquired strong momentum since the coming of age of new media such as Internet and the Web. They have become more relevant in a situation where growing communities and creating information categorizable through ontologies is made much easier and faster compared to what was possible before. In spite of this concomitance, the roles they have played in this information-rich environment have been so far not only different but also largely antithetic. The one played by communities is dynamic, and views information as something which is constantly changed and re-created by the agents that produce it. By contrast, the one played by ontologies views information in terms of its management at the meta-level through categories and concepts hierarchies, and it assumes that the ontology remains static, or changes very slowly as a consequence of decisions taken by the domain experts that control it. Given that information change is generally community-driven and this brings the clear necessity to make communities and ontologies interact. We propose to pursue this goal through a knowledge management approach, where the interaction between communities and ontologies is implemented as a knowledge life-cycle that leads to the creation of new concepts in the ontology as a consequence of the evolution of the information spaces constantly extended and re-created by the communities.

1 INTRODUCTION

Ontologies have since a long time provided a powerful tool to organize knowledge. At a philosophical level, ontology is the most fundamental branch of metaphisics. It studies being or existence and its basic categories and relationships, to determine what entities and what type of entities exist. At the more specific level of knowledge representation and knowledge management, ontologies identify concepts applied to specific domains and organized as graphs via relationship links. A typical example of an ontology as shown in Figure 1 is given by an automotive ontology, organizing concepts used by enterprises operating in the automotive industry.
Domain ontologies are traditionally the product of *panels*, *teams* and *committees* of domain experts and knowledge engineers. As such they are designed, maintained and evolved by these organized groups on the basis of the needs and objectives of the larger organizations they belong to. However, in a situation where organizations and corporations act less and less as the closed information sylos of the industrial age and are indeed compelled to re-act and co-act with an information-rich environment in order to prosper and survive, this approach appears too rigid and static. The desideratum would rather be one where the *communities* that provide the user/stakeholder bases for the products and services of organizations give also the input for the evolution of their conceptual infrastructure, so as to effectively capture and reflect dynamically the evolution of user needs and market trends. Domain experts and knowledge engineers would still be involved, but in an effort of combining and rationalizing knowledge effectively emerged from the bottom, rather than of imposing concepts more or less arbitrarily decided at the top. Thus, the ultimate goal is to make perfectly synchronous the alignment between organizations and their user and stakeholder communities, and to fully exploit the enormous potential for concepts

Arcelli Fontana F., Formato F. and Pareschi R..
ONTOLOGIES AND COMMUNITIES CO-EVOLUTION IN INFORMATION SYSTEMS .
DOI: 10.5220/0003106404530458
In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development* (KEOD-2010), pages 453-458
ISBN: 978-989-8425-29-4

creation deriving from such information sources as the new digital media, Web and the Internet *in primis*.

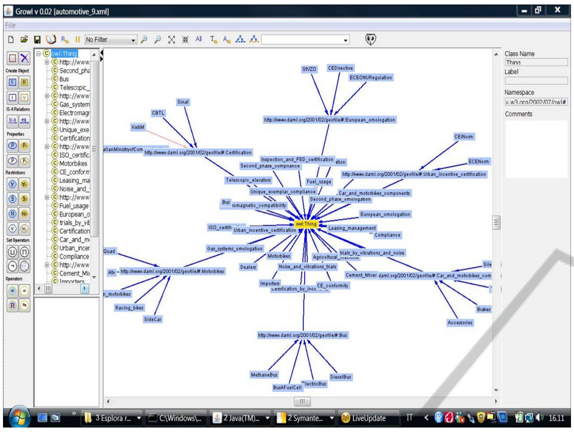

Figure 1: AN automotive ontology.

As we aim to show here, achieving this goal has some interesting consequences at a general foundational level, since it implies reconciling the “simple” and the “complex”, by making rational design and planning interact with the turbinous growth patterns of real life (even when in the form of “digital” life). Indeed it turns out that this is obtained by combining in a consistent way two different scientific traditions: one, rooted in philosophical logic and knowledge representation, and concerned indeed with the “simple” and the “rational” — namely, the study of ontologies and of their logics; the other, rooted in network theory, focused on the “complex” and the “emergent” — namely, the study of communities and of social networks.

The theoretical building blocks of our approach can be described as follows:

1. An “object” level where communities create content and information, and a “meta-level” where content and information is classified into ontologies;
2. A “knowledge life-cycle” that defines the interaction between the two levels in 1. (that’s where the main novelty of our approach lies) and makes possible to up-raise the process of information-creation at the object level into a process of knowledge-creation at the meta-level through the introduction of emergent concepts — concepts that, once certified and stabilized by teams of experts, can flow back into communities where they are adopted and shared.

1.1 Related Works

The problem of ontology evolution has been addressed by several authors in the literature. For example in (Stojanovic et al., 2002), the author claim that since generally ontologies grow in size, this requires a well structured ontology evolution process and they introduce the concept of an evolution strategy encapsulating policy for evolution with respect to user’s requirements.

In (Noy and Klein, 2004), the authors, in the context of ontology-evolution frameworks, analyze the similarities between database-schema evolution and ontology evolution which allow to develop an extensive research in schema evolution. In (Klein and Noy, 2003), the authors address the importance of ontology evolution in distributed development and they present an ontology of change operations, which is the kernel of a framework they proposed.

For what concerns with our research, in a previous work (Arcelli et al-a, 2009) we define a model by which ontologies evolve through Web community extraction. While in another work (Arcelli et al-b, 2009) we have introduced a methodology based on complex network parametrization, that studies the evolution of complex networks through an operator on graphs, whose purpose is to equalize meta-ontologies in the model we have proposed. Here, we describe an approach, with a wide scope both in terms of foundations and applications, that is based on the techniques and the apparatus we have described in our previous works. Hence our contribution here comes in the form of a research manifesto.

2 REPRESENTING ONTOLOGIES AND COMMUNITIES

Our view both of ontologies and of communities is information-driven: they are identified with the information they contain, either because they produce it (in the case of communities) or because they categorize it (in the case of ontologies). Furthermore, both communities and ontologies can be represented as networks (directed graphs). This common formal representation makes it easy to model the interaction between the two levels, yet it does not hinder us from identifying specific topological properties of the different types of networks that will be used to represent, respectively, ontologies and communities.

Indeed, ontology networks are typically characterized by a fixed number of nodes, corresponding to concepts, connected via a uniform distribution of links. This follows from their nature of networks planned and designed in a controlled fashion, with the aim of providing a complete and consistent conceptualization of a certain piece of knowledge.

By contrast, community networks are typically characterized by such phenomena as preferential attachments, meaning the fact that some nodes will be pointed to more than others as a consequence of their role of “hubs” and “leaders”. Furthermore, community networks will grow dynamically as more members join the community. As a very important *caveat*, it should be made clear that the notion of community that we adopt here not only assumes networks as a form of representation, but is itself a specialization of the notion of network: in fact, we adhere to the view, coming from the tradition of network theory, that a community can be defined in topological terms as a region of a dynamic network where links are denser than in the surrounding regions. In other words, communities are directly identified with highly interconnected regions of dynamic networks, as shown in Figure 2. This allows us to model as communities social networks whose nodes map directly into human individuals, such as family clans, but also digital communities where the role of humans is crucial but indirect, in that the primary community members are Web sites pointing one to the other. As we shall discuss later on, the most immediate applications for our approach to co-evolution of communities and ontologies are indeed in the domain of this kind of Web communities.

From the formal standpoint of network theory, the uniform link distribution of ontologies corresponds to networks-as-lattices as utilized in knowledge representation, see for instance the classical book by (Sowa, 1999) for a general overview of the subject.

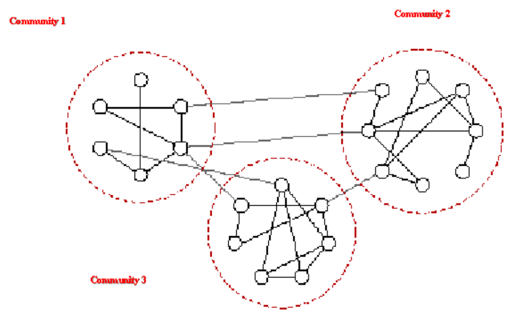

Figure 2: Communities within a network.

The preferential attachment behavior of community networks is formally accounted for by the scale-free networks recently studied by Barabási and his associates (Albert and Barabasi, 2002).

3 KNOWLWDGE LIFE CYCLE

Precisely because we view the world of organizations, expert teams and ontologies on one side, and the world of communities and emerging concepts and experiences on the other, as communicating rather than as separated, we aim to define a model through which they can fruitfully interact, thus making possible the circulation and re-creation of conceptual knowledge. This healthy circulation is the opposite of the corporate ailment of knowledge stagnation, which takes place whenever organizations lose touch with the needs and feelings driving the communities of the stakeholders they depend on — perhaps the most deadly of corporate disesases in an era where enterprises can effectively compete on communication and knowledge transfer.

The model of lifecycle that we adopt is itself an adaptation of the well-known “Double-loop Learning” model developed by (Argyris and Schön, 1978), which has found vast and effective application in the management of many types of knowledge processes in a variety of organizations. We specifically apply it to the interaction between organizations and the social networks of stakeholders existing at their borderline — a phenomenon which has emerged forcefully with the extended communication spaces of the new millenium.

Argyris and Schön distinguish between single-loop and double-loop learning, related to Gregory Bateson’s concepts (Bateson, 1979) of first and second order learning. In single-loop learning, individuals, groups, or organizations modify their actions according to the difference between expected and obtained outcomes. In double-loop learning, the entities (individuals, groups or organization) question the values, assumptions and policies that led to the actions in the first place; if they are able to view and modify those, then second-order or double-loop learning has taken place. Double- loop learning is the learning about single-loop learning.

4 ONTOLOGY (CO)-EVOLUTION

Quite obviously, single-loop learning takes place in applying an ontology to the domains it is supposed to categorize. Take for instance a wine ontology and consider a directory of Web sites related to wine, including wine sellers, wineries, wine clubs etc. Then one simple procedure to learn how to use the ontology is as follows:

- we select Web sites from the directory and we associate them with nodes in the ontology;
- the concepts in ontology sites get "trained" with the content in the Web sites, through some classifier algorithm such as bayesian inference, neural networks, or support vector machines;
- by following the links of the Web sites, we apply the trained concepts to new content:
  - if they classify according to expectation then we have reached the appropriate training of the ontology and thus we have learned how to use it;
  - Otherwise we might need to refine their training so as to effectively make them capable to clasify Web sites as expected.

Now, what about double-loop learning? According to the definition, this must question the structure of the ontology itself by bringing as a consequence the introduction of altogether new concepts. As a matter of fact, this is what happens whenever teams of experts revise ontology structures in order to adapt them to changes in industry and market trends. Thus, such changes are generally re-active to mutated conditions in the environment; therefore, what is still missing is a sound methodology to alert the experts of the need of change and drive them in the right direction. In order to answer to this need, we apply double-loop learning as follows:

- let's assume that we have applied single-loop learning as above. This means that we have partitioned a part of the Web into "concept graphs", identified by the scope of application of the concepts in the ontology;
- then let us explore this portion of the Web watching for communities (namely, highly interconnected regions) and distinguishing two cases:
  - communities corresponding to existing concept graphs;
  - communities that do not fit with existing concept graphs (even if they may be partially overlapping with existing graphs)
- if the second such case occurs, then this is an indication that we are in front of one or more new concepts, and that the overall current architecture of the ontology must be revised;
- it is then the work of the experts to acquire this input and elaborate it through the various techniques available, starting from the inspection and the analysis of the Web sites belonging to the uncovered communities, and to extend and revise the ontology accordingly.

Both these steps are summarized in Figure 3, which depicts double-loop learning as applied to ontology/community co-evolution.

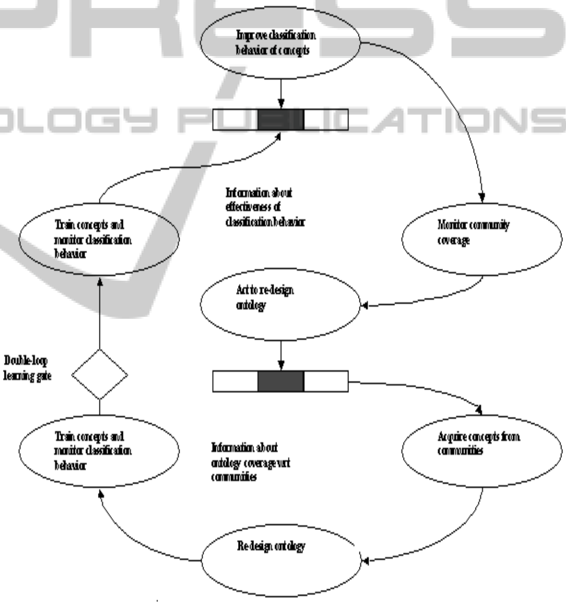

Figure 3: A double-loop learning system for community/ontology co-evolution.

As an example, suppose that we find a community of Web sites which is not covered by any existing concept in the corresponding wine ontology. On the other hand, a sub-region of this community is indeed covered by a concept graph corresponding to the concept of "White Zinfandel" (the rosè wine from California). The remaining part of the community is characterized by content related to Italian ham and salami. On the basis of further content analysis, this may be taken as an indication that White Zinfandel lovers see Italian ham and salami as a suitable food match for their favourite wine, and that creating a corresponding "menu concept" may be relevant and approparite, with possible applications to the design of new products packaging White Zinfandel and Italian ham and salami to be distributed in wine shops, food stores and shopping malls.

It should be pointed out that, beside double-loop learning, this approach could be viewed as fitting within other knowledge management methodologies such as, in particular (Nonaka and Takeuchi, 1995) Knowledge Spiral, which can be considered itself as evolving further the concept of double-loop learning. The Knowledge Spiral defines a cycle of four phases, given by Knowledge Internalization, Knowledge Socialization, Knowledge Externalization, and Knowledge Combination. In our context, Knowledge Internalization and Knowledge Socialization play the role of single-loop learning as viewed above, namely as learning to apply the ontology to the relevant part of the Web, while Knowledge Externalization and Knowledge Combination play the role of double-loop learning in the different phases of identification of a new concept from the Web and of consequent re-design of the ontology.

Finally, the strong use of tools from Information Technology and Artificial Intelligence to support the automation of the different phases of the learning cycle, such as content classification and analysis algorithms, suggests also, for the purpose of its general support, the definition of a Knowledge Management IT Architecture in the sense of (Borghoff and Pareschi, 1998).

5 APPLICATIONS

The idea of leveraging in a systematic way the eco-system that connects organizations with their surrounding communities in order to pursue concept creation has a very wide potential, with applications that, in different ways, reach the very core issues of innovative design of products and services. Here we highlight briefly two specific domains that appear as particularly relevant: user-driven innovation and a "community-oriented" version of the Semantic Web project.

User-driven Innovation. There are two main approaches to product innovation. In the so-called "linear model" the traditionally recognized source is manufacturer innovation. This is where an agent (person or business) innovates in order to sell the innovation. Another source of innovation, only now becoming widely recognized, is end-user innovation. This is where an agent (person or company) develops an innovation for their own (personal or in-house) use because existing products do not meet their needs. Eric von Hippel has identified end-user innovation as, by far, the most important and critical in his classic book on the subject, *The Sources of Innovation* (von Hippel, 1988). One outstanding example of end-user innovation is open-source and free software.

However, while many users may correctly identify the need of innovation, they may lack the technical skills or the economical means or simply the will to innovate. Ideally, this situation could offer excellent opportunities for manufacturers to innovate effectively, if the could listen carefully enough to their user communities, thus providing an intermediate model between manufacturer and end-user innovation. Of course, this idea is not new but so far it has not been obvious how to put it in practice. User groups as supported and implemented nowadays by many enterprises go in this direction, but, again, they imply the willingness of users to organize themselves in somewhat formal structures, which may be less productive and creative with respect to the totally free format given by communities.

Double-loop learning to make innovative concepts emerge from communities and enter, with an effect of creative disruption, corporate ontologies may provide an important basis to evolve this potential for product innovation into a fully practicable methodology.

Semantic Web. The Semantic Web http://www.w3.org/2001/sw/ is a project, managed "from-the-top" by standard committees and research institutions, to make the Web fully "understandable". (For an overview, from the point of view of the Semantic Web founders, of where the Semantic Web stands since its inception in the very early years of this millenium see (Berners-Lee et al., 2006). In this way, software agents could inspect content of the Web pages and automate e-business and e-commerce actions. On the other hand, by moving from the "bottom" ground of people and communities, the primeval Web (so called Web 1.0) has evolved on its own into something completely different, Web 2.0 — namely the Web of blogs, social networks and personal spaces. There is a general consensus that Web 2.0 is, first and foremost, about people, and is neutral and open to any kind of technologies or standards as long as they provide support to people-oriented applications.

We view our approach as instrumental to reconciling the quest of semantic clarity initiated by the Semantic Web project with the explosive growth of people-oriented Web 2.0. The point is that the Semantic Web, as originally conceived, implies a strong management of the information available on Web sites, by annotating it manually with semantic meta-information such as XML tags, ontologies and "resource-description frameworks". This contrasts with the way people use the Web, and Web 2.0 in particular, that is essentially for communication and personal networking, caring a lot about content and not too much about meta-content. By providing a way to co-evolve communities and ontologies, our framework can be exploited to automate the creation of a "meta-web" where the burden of semantic annotations is taken away from the users.

6 CONCLUSIONS

As in all the research manifestos, the conclusion is the beginning of… the beginning, and the main thing that can be said is that we expect much from what has to come. But, just to summarise a bit, we state again the goal of this research program: which is of viewing communities and ontologies, two concepts that have both gained strong momentum through the coming of age of the new media, as fully complementary even if they move from apparently distant premises, viz. emergent behavior in one case and rational design and planning in the other. The result shall be a novel knowledge lifecycle aimed at avoiding knowledge stagnation through the constant generation of fresh concepts, and the consequent re-design of the ontologies that host them — a result obtained by combining in a non-intrusive way the creative force of communities and the rational design of knowledge teams.

REFERENCES

Albert, R., and Barabási, A., 2002. Statistical mechanics of complex networks. *Review Modern Physics*, 74, 47.

Arcelli, F. Formato, F. and Pareschi, R., 2009a. Ontology Engineering: Co-evolution of Complex Networks with Ontologies**,** *Proceedings of the Workshop on Ontologies for e-Tchnology* (OET 2009), Italy.

Arcelli, F. Formato, F. and Pareschi, R. 2009b. Equalizing the structures of web communities in ontology development tools. *Proceedings of the International Conference on Intelligent Systems Design and Applications*, (ISDA'09), Italy.

Argyris, C., and Schon, D., 1978. Organisational learning: A theory of action perspective. Addison Wesley, Reading, MA.

T. Berners Lee, N. Shadboldt and W. Hall, 2006. The Semantic Web Revisited. *IEEE Intelligent Systems*, 21 (3).

Bateson, G., 1979. Mind and Nature: A Necessary Unity, Advances in Systems Theory, Complexity, and the Human Sciences. Hampton Press. ISBN 1-57273-434-5.

Borghoff, U. M., and Pareschi, R., 1998. Information Technology for Knowledge Management. Springer-Verlag, Berlin and Heidelberg, Germany

von Hippel, E., 1988. The Sources of Innovation. Oxford University Press, Oxford, UK.

Klein and Noy, 2003. A Component-Based Framework For Ontology Evolution. *Proceedings of the IJCAI Workshop on Ontologies and Information Sharing*, Seattle, WA.

Nonaka, I., and Takeuchi, H. (1995) The Knowledge Creating Company. *Oxford University Press*, Oxford, UK.

Noy and Klein, 2004. Ontology Evolution: Not the Same as Schema Evolution, Journal of Knowledge and Information Systems, Vol.6, N.4, Springer London

Sowa, J. F., 1999. Knowledge Representation: Logical, Philosophical, and Computational Foundations. Brooks Cole Publishing Co., Pacific Grove, CA.

Stojanovic et al., 2002. User-Driven Ontology Evolution Management, Knowledge engineering and knowledge management: Ontologies and the semantic web, LNCS 2473/2002, Springer.