

Misinforming Knowledge through Ontology

Bill Davey
RMIT, Melbourne,
Australia

bill.davey@rmit.edu.au

Arthur Tatnall
Victoria University, Melbourne,
Australia

arthur.tatnall@vu.edu.au

Abstract

A fundamental of the formation of virtual organizations, or capturing effective organizational memory is the creation of an ontology. Existing organizational memory software systems uniformly suffer from the problem of initializing the ontology. Another, less recognized, problem of ontologies is their poor record in capturing implicit and informal relationships within and between actors. This paper postulates that previous work on an ecological model of relationships can be applied to bootstrapping an ontology. This model seeks to capture the complexity of relationships using language arising from the areas of ecology which has a long history of dealing with complex relationships.

Keywords: Knowledge management, ontology, bootstrapping, virtual organisations, ecological models

Introduction

The World Wide Web Consortium (W3C) has now a recognised ontology modelling language (W3C 2004). Tools exist such as Chimera from Stanford University (Stanford, 2004) that allow us to [perform analysis on ontologies embedded in web sites. These developments allow us to use software agents to deliver meaning rather than flat text. This meaning can come from the web, or from documents within an organisation, analysed through the organization ontology. When we rely on computer systems to support decision making with derived knowledge, the underlying model used by the computer system must be trusted. Systems based on ontology have a serious flaw in that the ontology grows from a seed that defines the breadth of knowledge that can be organised. If this seed (called the bootstrap) includes domain knowledge, but no knowledge of the people working in the domain, then we cannot trust results.

Webster's Dictionary defines ontology as "the branch of metaphysics dealing with the nature of being, reality, or ultimate substance." However, Abou-Zeid (2003) recognises that, for the AI (artificial intelligence) community what 'exists' is that which can be represented, and ontology from this perspective is "an explicit specification of conceptualization." Knowledge management systems can 'create' knowledge by semantic analysis of a database (often called an 'organisational memory') of data gathered from interactions between members of an organisation and the knowl-

edge management system. To perform such a semantic analysis the knowledge management system must have created a formal language specification of all the concepts used in the organisation. This formal specification of 'what is' for an organisation is the ontology. An illustration taken from

Material published as part of this journal, either on-line or in print, is copyrighted by Informing Science. Permission to make digital or paper copy of part or all of these works for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage AND that copies 1) bear this notice in full and 2) give the full citation on the first page. It is permissible to abstract these works so long as credit is given. To copy in all other cases or to republish or to post on a server or to redistribute to lists requires specific permission from the publisher at Publisher@InformingScience.org

software of Sheth, Thacker and Patel (2003) shows this:

```
Earthquake (latitude, longitude, region,  
event Date, description,  
damagePhoto,  
numberOfDeaths, magnitude);
```

This definition is of a class that will define the ontological entity earthquake. In this problem domain, whenever the term ‘earthquake’ is used it must have the parameters: latitude, longitude, region etc and nothing outside this list of parameters has anything to do with earthquakes. By tightly specifying the language of a problem domain we allow the knowledge management system to perform analysis amongst the data captured.

Software to date has been capable of gathering fairly mundane engineering aspects of the processes and operations of organizations. Current research into knowledge acquisition focuses on enriching knowledge capture and creation, and this enrichment requires more sophisticated ontology to underlie the process. What can happen if the ontology is flawed is that analysis is based on oversimplified concepts leading to conclusions that either have little relevance or are flawed as they leave out important factors. This is particularly important when relationships are omitted from an ontology. Imagine a virtual organisation containing two companies; a wine maker in the Barossa Valley and a wine distributor with warehousing in Melbourne and Brisbane. Analysis of captured data from a time and motion study is in the form “shipping to Melbourne will take 24 hours, shipping to Brisbane will take 36 hours” and this is used to make scheduling decisions. In operation the system is found to deliver to Melbourne in 24 hours and to Brisbane in 52 hours. Detailed study of the delivery shows that the trucks going to Brisbane are always loaded last at the winery because a competing company shipping to Brisbane pays more for the wine and is given preference. This class of problem with ontology is due to the difficulty of starting an ontology. Once seeded with useful data an ontology will grow to include all, important factors. The problem is that seeding or ‘bootstrapping’ of ontologies often ignores whole areas of data, particularly that data to do with relationships between actors. In this paper we draw together research from the fields of actor network theory, ecology, management theory and ontology to point to a solution for a common problem with ontology formation.

Creating New Knowledge

Although the idea of capturing domain concepts in machine terms may be interesting, the field of ontology has become very useful in real organizations. Alani, Kalfoglou, O’Hara and Shadbolt (2002) exemplify researchers questing for higher levels of analysis of the concepts captured in machine language. They claim that “the advantage of studying ontologies is that the concepts and relations therein have explicit semantics or types, and therefore providing another source of information over and above connectivity or simple subsumption. This semantic information can be taken account of when performing a network analysis, allowing raw results to be refined on a relatively principled basis.”(Alani, Kalfoglou, et al. 2002) These researchers, and others, have created systems that find new connections between concepts and new relationships between people and hence create new organizational knowledge. One line of use of these systems that create knowledge from semantic analysis is that of organizational memory systems. These systems seek to monitor operations in an organization, capturing new entries for the ontology and adding each entry to the material for semantic analysis. These systems are software that make explicit Nonaka’s (1994) knowledge conversion process which recognises that while explicit knowledge is the knowledge that can be articulated in communicable forms, tacit knowledge is the personal knowledge embedded in individual experience and involving such intangible factors as personal

beliefs, perspectives, and values. The interaction between these two types of knowledge takes place in four basic processes of knowledge conversion.

The Problem of Bootstrapping

There are two problems in the dream of systems automatically finding knowledge by tapping into the operations of an organization. Firstly the problem of bootstrapping. Alani, Kalfoglou, et al (2002) have discovered the bootstrapping problem: that the system must be used before it has any interesting content, but that precludes its having any interesting content in its initial conditions. Secondly, the current computer languages that seek to represent knowledge have come from an engineering perspective and lack the ability to capture all classes of knowledge easily. Sheth, Thacker and Patel (2003) claim that relationships between entities (terms or concepts) are the basis of capturing, representing, and supporting semantics. Present-day ontology languages, such as those based on RDF/RDFS, F-Logic, Conceptual Graphs, SHOE, Description Logic (including DAML+OIL) provide formal representations of relationships and the ability to have relatively efficient inferencing capabilities. However, the complex, user-defined, inter-ontological relationships that are required to deal with complex relationships are beyond the expressive and computational capability of languages like the ones listed above. In fact the practitioners of ontology research have come to recognize that useful ontology creation is not amenable to blind analysis. Perez (1999) describes ontological engineering as “still a craft rather than an engineering activity”.

Techniques that show this blind ‘can’t think of anything better’ approach abound. Alani, O’Hara, and & Shadbolt, (2002) describes the process used in one major system: “Upon start-up, an ontology gives Quickstep an initial set of publications for each user. Each user’s known publications are correlated with Quickstep’s paper database, and a set of interests is then generated for the users. This provides one way of overcoming the initialisation problem for the new system.” ONTOCOPI (Alani’s system name) uses a convenient source of data (the publications of a user) in an arbitrary way to seed the ontology. Areas of research that have experience in capturing data in complex situations offer an alternative to this random seeding process.

Research over the last few years (Tatnall, 2000, 2002; Tatnall, & Davey, 2001, 2002a, 2002b), Tatnall and Davey (2003) have shown that concepts from ecology are capable of helping us understand dynamic relationships between actors in an organization. If these fundamental concepts are made part of the initial state of an ontology then a proven framework for understanding the relationships exists in the system. By using this proven approach we can help overcome the problem commonly expressed by makers of organisational memory systems. This problem has been identified in field surveys (Dieng, Corby, Giboin, & Ribiere, 1999) as well as in implemented systems (e.g. Althoff, Nick, & Tautz, 1999; Alani, Kalfoglou, et al. 2002).

Ecology

Ecology is a relatively new science, being proposed in 1869 by Ernst Haeckel as “the total relations of the animal to both its organic and its inorganic environment” (Krebs, 2001, p. 2). The dictionary defines ecology as: “the branch of biology dealing with the relations of organisms to one another and to their physical surroundings” (Pearsall & Trumble, 1996, p. 446). Ecology is related to other biological studies as shown in Figure 1 below.

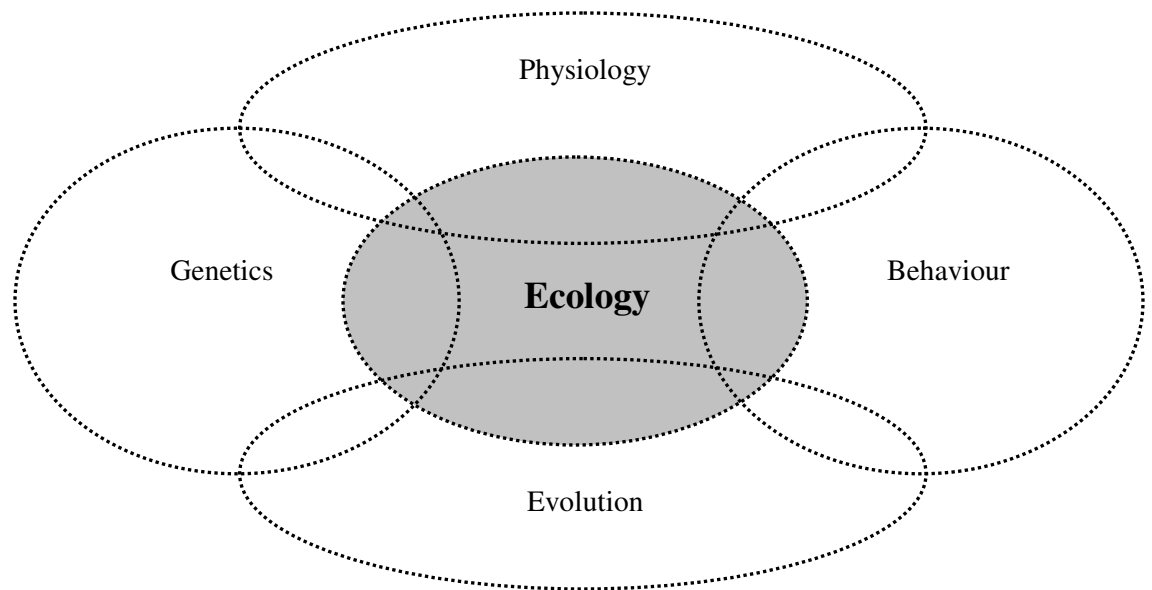


Figure 1: Ecology and other branches of biology (- adapted from Krebs, 2001, p. 2))

Ecology is concerned with interrelationships; between different living things, and between living things and their environment. More specifically, Krebs (2001 :2) defines ecology as “the scientific study of interactions that determine the distribution and abundance of organisms.” It is thus concerned with the way that organisms respond to the various forces that operate within the environment.

Two key biological principles exemplify the concepts of ecology (Townsend, Harper & Begon 2000): organisms behave in ways that optimise the balance between their energy expenditure and the satisfaction they obtain, and organisms operate within a competitive environment that ensures only the most efficient of them will survive.

Habitat, ecological niches, the exploitation of resources in predator-prey interactions, competition, and multi-species communities (Case 2000; Krebs 2001) are all important considerations in ecology. Many different individuals and species typically occupy any given ecosystem, and they can be considered to interact in many different ways. An ecosystem is a highly complex entity due to the large number of living things inhabiting it, and to the variety of interactions possible between each of these (Tatnall & Davey, 2002a). The ‘ecosystem’ represented by an organization would contain, at least, human ‘species’: and many inanimate objects. The ecological model suggests that if interactions between actors (people, organizations, processes, systems) will tend to maximise output for energy input. An ontology engine should look at the satisfaction obtained from energy expenditure.

In an ecosystem many different individuals and species typically occupy a similar space and can be considered to interact in the following ways:

- **Competition.** In biology this occurs when two individuals or species each strive for the same thing, which is typically food, space or some other physical need. Competition in nature can occur both within and between species. Organization systems contain many examples of competition. Sometimes this can be seen between two workers with different philosophies or approaches. This can lead to various problems including time-wasting clashes of personality between workers. Competition may also be perceived in the form

of scarce resources being use up by a competitive technology, or in determining the ‘fit test’ systems and techniques best suited for survival (Darwin 1958) in the curriculum.

- **Co-operation.** When one population is benefited by the presence of another, but can survive in its absence we have a co-operative situation. There are several variants on co-operation in ecology depending on whether each entity benefits or whether there is a degree of harm to one entity involved.
- **Filling a niche,** often by occupying an unfavourable location, is a technique used by some species to avoid competition.

Ontology and Ecology

Although not seen in a coherent model, these concepts have been found by many authors, particularly those looking at virtual organisations, where naturally competing interests must make way for co-operative ventures. El-Sayed Abou-Zeid (2003) found it useful to think in terms of that “cells” co-operating with other specialised units and that most current techniques overlook social aspects in their steps. We postulate here that an ontology can be seeded from known relationships between competing and co-operating entities, hence involving relationships from the very start of an ontology creation. The OWL definition of an ontology allows three classes (W3C2004):

- Classes (general things) in the many domains of interest
- The relationships that can exist among things
- The properties (or attributes) those things may have

What we are suggesting here is that a new set of “things” – the relationships competitor, co-operator and energy efficiency, should be included in all ontologies. In terms of the W3C definition this makes a type of “meta relationship” which captures information about the underlying relationships that we postulate exist in all circumstances.

Misinforming Ontology

What happens if an ontology is not bootstrapped usefully? Our previous studies have indicated that knowledge of relationships between actors is the most useful explanation system in the domain of innovation translation. years (Tatnall, 2000, 2002; Tatnall & Davey, 2001, 2002a, 2002b, 2003). If an ontology is seeded with elements that reflect only known technical terms from the domain, it cannot grow in unexpected directions. An ontology that ignores the elements of a domain that are most useful for explanation will still produce knowledge (conclusions) from the semantic analysis, these will just be wrong. Current ontology models are incestuous. Cranefield, Haustein, & Purvis (2001) describe this as “These systems support the definition of concepts by simply naming them and specifying where they fit in the generalisation/specialisation hierarchy of existing concepts.” An ontology that is essentially self referential can seriously misinform us as to the knowledge produced by analysis.

Conclusion

The potential for basing analytical techniques on sound ontologies is immense. Researchers in this field know that the challenge is to find ways of modelling real life more accurately and richly. Alani, Kalfoglou, et al. (2002) warn that many points of interest in an organization’s ontology will not be spotted by the methods involved, especially if the ontology is in some way incomplete, and fails to cover the object domain fully in some important respect.

We can see here how the application of proven ecological models can both address the problem of finding initial relationships between actors and of bootstrapping a system with information known to be relevant.

A system can only perform useful semantic analysis if a large amount of data has been gathered, leading to the likelihood of correlations between ontological entities. An ontology that is seeded with well known relationships will be more immediately useful; to employees. This makes the data gathering that leads to valuable knowledge creation much more likely.

References

- Abou-Zeid, E. (2003). What can ontologists learn from knowledge management? *The Journal of Computer Information Systems*, 43 (3), 109-118.
- Alani, H., Kalfoglou, Y., O'Hara, K. & Shadbolt, N. (2002). Initiating organizational memories using ontology-based network analysis as a bootstrapping tool. *Expert Update*, 5(3), 43-53.
- Alani, H., O'Hara, K., & Shadbolt, N. (2002). ONTOCOPI: Methods and tools for identifying communities of practice. *Proceedings of the 2002 IFIP World Computer Congress*, Montreal, Canada, 2002a.
- Althoff, K-D., Nick, M., & Tautz, C. (1999). Improving organizational memories through user feedback. In F. Bomarius (Ed.), *Proceedings of the Learning Software Organizations (LSO'99) Workshop (27-44)*. Kaiserslauten, Germany.
- Case, T. J. (2000). *An illustrated guide to theoretical ecology*. New York: Oxford University Press.
- Cranefield, S., Haustein, S., & Purvis, M. (2001). UML-based ontology modelling for software agents. In *The Information Science Discussion Paper Series Number 2001/07* May 2001
- Darwin, C. (1958). *The origin of species*. New York: The New American Library. (Original work published in 1859.)
- Dieng, R., Corby, O., Giboin, A., & Ribiere, M. (1999). Methods and tools for corporate knowledge management. *International Journal of Human-Computer Studies (IJHCS)*, 51, 567-598.
- Krebs, C. J. (2001). *Ecology - The experimental analysis of distribution and abundance*. San Francisco, Benjamin Cummings.
- Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organizational Science*, 5 (1), 14-37.
- Pearsall, J. & Trumble, B. (Eds.). (1996). *The Oxford English reference dictionary*. Oxford: Oxford University Press.
- Sheth, A, Thacker, S., & Patel, S. (2003). Complex relationships and knowledge discovery support in the InfoQuilt system. *The VLDB Journal*, 12, 2.
- Stanford University. (2004). Chimæra. Retrieved March 2004 from <http://www.ksl.stanford.edu/software/chimaera/>
- Tatnall, A. (2000). Innovation and change in the information systems curriculum of an Australian University: A socio-technical perspective. PhD thesis. Education. Rockhampton, Central Queensland University.
- Tatnall, A. (2002). Modelling technological change in small business: Two approaches to theorising innovation. In S. Burgess (Ed.), *Managing information technology in small business: Challenges and solutions* (pp. 83-97). Hershey, PA: Idea Group Publishing.
- Tatnall, A. & Davey, B. (2001). How Visual Basic entered the curriculum at an Australian University: An account informed by innovation translation. *Proceedings of the Informing Science Conference*, Krakow, Poland.

- Tatnall, A. & Davey, B. (2002a). Information systems curriculum development as an ecological process. In E. Cohen (Ed.), *IT education: Challenges for the 21st century* (pp. 206-221). Hershey, PA: Idea Group Publishing.
- Tatnall, A. & Davey, B. (2002b). Understanding the process of information systems and ICT curriculum development: Three models. In K. Brunnstein & J. Berleur, (Eds.), *Human choice and computers: Issues of choice and quality of life in the information society* (pp. 275-282). Assinippi Park, MA: Kluwer Academic Publishers / IFIP.
- Tatnall, A. & Davey, B. (2003). ICT and training: A proposal for an ecological model of innovation. *Educational Technology & Society*, 6 (1), 14-17.
- Townsend, C. R., Harper, J. L. & Begon, M. (2000). *Essentials of ecology*. Massachusetts, Blackwell Science.
- W3C. (2004). OWL Web ontology language overview. W3C Recommendation 10 February 2004

Biography

Bill Davey is a Senior Lecturer in the School of Business Information Technology at RMIT University, Melbourne, Australia. He holds bachelors degrees in Science and Education, a Graduate Diploma in Computing, and Master of Business degree. His research interests include methodologies for systems analysis and systems development, Visual Basic programming, information systems curriculum, and information technology in educational management.

Arthur Tatnall is a Senior Lecturer in the Victoria Graduate School of Business at Victoria University in Melbourne, Australia. He holds bachelors degrees in Science and Education, a Graduate Diploma in Computer Science, and a research Master of Arts in which he explored the origins of Information Systems curriculum in Australian universities. His PhD involved a study in curriculum innovation in which he investigated the manner in which Visual Basic entered the curriculum of an Australian university. His research interests include technological innovation, information systems curriculum, Visual Basic programming, project management, electronic commerce, and information technology in educational management.

Arthur and Bill have worked together and co-operated on many occasions. They have worked on several joint research projects and co-authored a large number of papers, book chapters, and textbooks relating to management information systems, programming and IS curriculum.