What are we Modeling when we Model Knowledge?

Steve McKinlay
School of Information Technology, Wellington Institute of Technology,
Petone, New Zealand.
steve.mckinlay@weltec.ac.nz

This paper examines the relationships between natural language and our attempts to codify and manage business knowledge in particular the normative statements that spawn all business decisions. I argue that such knowledge is inextricably bound to language and experience and that our conceptualizations, which ultimately constitute knowledge, are inseparable from language. Furthermore if we wish to develop a model of knowledge that can be represented computationally we need to understand the extent to which language can be mapped to its empirical content and or to sentences, which correlate with equivalent sentences. I argue that under current computational frameworks such a project is doomed due to inherent indeterminacies in terms and the concepts those terms seek to denote. I conclude thus that knowledge cannot be logically and systematically represented under our current and traditional computational frameworks[1].

Keywords

Knowledge, Data Modeling, Natural Language, Indeterminacy, Knowledge management.

1. INTRODUCTION

One might ponder the worth in being concerned with clarifying the logical nature of normative statements[2]. The kind of statements I refer to are those which we might consider to be, or at least be a part of *business knowledge*. Normative statements are the kinds of statements that help managers make decisions.

The modeling of such statements I suggest ought to be precisely the worry of those wishing to model knowledge for it is an organizations ability to materialize successful judgments in conjunction with relevant empirical data that characterize its knowledge or lack thereof. It follows that the success of decisions based on such normative statements ultimately determine the success of any organization and thus I venture its knowledge management system.

What is evident to any manager is that "true" and "false" are held largely inapplicable to such statements. *Belief* albeit distinct in nature seems as important as *knowledge*[3] because an organization must have the belief that they are making the right decision in order that some degree of success obtains out of the decision making process. Certainly the prospect of acting upon a normative business decision absent in the belief that it would be *good* for the business would not seem to bode well for the success of the decision.

Given the failure rate of IT based "knowledge management" projects it might be at least economically prudent if not philosophically interesting to investigate some of the reasons for failures. This paper seeks to examine the issues, particularly the philosophically interesting ones; I'll leave it to the accountants to count the cost.

In this paper I argue that modeling normative statements is tantamount to modeling language, experience and belief. Furthermore I argue that the conjunction of language and experience is what constitutes knowledge and that attempts to model knowledge under current computational frameworks, that is those currently used to formally represent data is a venture bound for failure.

2. POSITIVISM REVISITED

A skeptical stance in regard to the logical representation of knowledge is not new. The failed verificationist project of logical positivism stands as a sombre testament to those who tried before us. This project sought to explicate a set of logically verifiable principles or a complete theory that would stand as foundational in underpinning all scientific theory. It turned out for a variety of reasons that this was simply not possible. I suggest an obligatory study of this historical project be compulsory for all would be knowledge modelers.

Current efforts in the domain of knowledge modeling I argue focus on the wrong problems, there is a herd-like obsession in regard to what particular

kinds or categories of knowledge might be modeled, proponents feel obliged to reiterate Polyani's *tacit* and *explicit* knowledge distinction as the starting point for any and all discussions.

> *"A distinction between tacit and explicit knowledge is critical to understanding the working mechanisms of knowledge management."*
>
> *(Gupta, McDaniel., 2002)*

The tacit/explicit distinction is not the salient issue for those wishing to model knowledge. Initially discussed during developments in modern logic in the 30s and 40s serious issues arose during attempts to develop a logical syntax which might be used to represent the higher level theoretical language under which all science might be modeled. The basis for any scientific meta-language would have to account for a connection between theoretical terms and observational terms which carry the empirical content[4]. Issues of greater concern arise when attempting to logically analyze normative or value statements. Absolute value statements are devoid of any cognitive meaning by any empirical standards, are normative in nature and thus defy logical representation.

The corollary to these problems is that the translation and reduction of context specific value statements expressed as natural language into something, which can be modeled within a logical formal system seems like the project of logical positivism doomed to failure.

To date our typical attempts to model data and consequently information within computer systems involve the formalization of structures we characteristically know as entities or objects *(the structural aspect*), relationships between those objects (*the integrity aspect*) and a set of operators that enables users to access the information (*the manipulative aspect*). (Date, p 58, 2000).

Such constructions appropriately formalized and defined represent data axiomatically within computer systems (usually databases). These structures are necessarily deductive and hence tautological in nature. That is, the execution of queries over such objects are in fact *truth-valued expressions*. Interpretative conclusions based on the presentation of such data go on in the head of analysts and managers. There is no room for interpretation within the formal system.

With relational databases in particular there is an *isomorphic* relationship between the predicates that define entities and the meaning of the subsequent rows (or tuples if you prefer) contained therein. This is in fact a good thing; we want our databases to contain precisely defined predicates thus guaranteeing the truth of the propositions which represent valid extensions of those predicates. The significance of any truth-valued expression would be lost if the meaning of such data were not fixed.

Furthermore any connection between natural language, the predicates represented in a database or the structures defined within a formal (syntactical) language are at best contingent. We could just as easily define database table headers (predicates) as x, y, z or similar and some of us do. This is in keeping with the formality of the logic (although not advisable by database administrators). This approach differs fundamentally from the implicit appeals to intuition and experience present in the way in which we use and understand natural language. There *is* no such one to one correspondence (except in some rare cases) between natural language terms and things or the concepts those terms seek to denote.

The point of difference ought to be clear. Natural language is semantically rich, contextual, instructive, and often normative, thus logically imperfect and unsystematic. The practical difficulties in deriving a set of systematic and logical rules of syntax for natural language with respect to certain formalized languages are profound and ultimately futile.

With regard then to "knowledge management" and associated modeling efforts point toward an acute confusion between modeling statements, propositions or entities which exist within a formal system where the foundational principles are well established and thus are logically derivable *a priori,* contrasted with the synthetic and often normative nature natural language statements. The latter exhibit no such foundational principles from which any knowledge base might grow from. The latter comprises of highly contextualized knowledge embedded in a localized dialect, which rely largely upon prior experience to confirm or detract from their confirmation.

This discussion is not intended to support the idea that natural language has any kind of "mystical quality" absent from formalized languages. The point is to clarify the usage of formalized languages in contrast to natural language.

3. KNOWLEDGE AND NATURAL LANGUAGE

At this point the reader may well ponder what strength of connection between knowledge, normative

statements and the natural language they are expressed in I have assumed. Quine deliberates "Each of us learns his language from other people, through observable mouthing of words under conspicuously inter-subjective circumstances." (1960, p1). This equally applies to children learning their first language to the kind or jargon and business oriented terms which colour typical natural language exchanges in the organization.

Such language develops as an ordinary language of physical things in relation to non-verbal stimulation. The development of conceptualizations are thus inseparable from language since they relate to other concepts (expressed linguistically) and or to our ordinary language of physical things.

If we wish to improve our understanding of our conceptualizations of more complex matters it does no good to attempt to impose some sort of reduction on the language within which they are expressed, such understanding comes from the clarification of casual connections between physical things and other matters. (*ibid*, p3). In fact there is empirical evidence in psychology to suggest that we do not reduce concepts into constituent sets of propositions, if this is the case attempts to model such reductions in data or knowledge bases certainly wouldn't be modeling knowledge in the way that it is represented in our minds.

Nevertheless it is through language that we develop our knowledge of the world. All manner of complex conceptual and intricately related talk may well bear little relation to non-verbal stimulation however the empirical content of such talk must be sheeted back to our sense-interaction with our world.

If then we wish to develop some kind of knowledge model we may wish to consider the extent to which language can be mapped to empirical content or the sentences in the language, which correlate with equivalent sentences.

Quine discusses such problems in detail. In his 1960 publication Word and Object Quine outlines what is termed the indeterminacy argument, summarized succinctly by William Alston (1986), that no one means anything determinate either by any of his terms or by any of his non-observation sentences. Given that the constitution of all language is derived from the speech activity of its users it follows that no term or non-observation sentence in a language means anything determinate.

Although on a first reading seemingly extreme, these claims ought not to be of surprise to anyone, semantic indeterminacy can be recognized (in its most simplistic form) as vagueness of degree, how many inhabitants does it require for a 'village' to be called a 'town' or a 'town' a 'city'? Indeterminacy in meaning is well documented by epistemologists and logicians alike, Wittgenstein's "family resemblance" (1953), Waismann's "open texture" (1945) and what Michael Dummett (1974) termed the "inextricability thesis", "the view that there is no sharp line between what belongs to the meaning of a word and what belongs to widely shared and firmly held beliefs about what the word denotes". (Alston, 1986)

4. IMPLICATIONS FOR KNOWLEDGE MANAGEMENT

If there is no objectively correct translation of individual terms then the complexities associated with regard to codifying multifaceted business value statements and their relationship with particular users experience, intuition or states of belief are profound.

The point to be made is the significance in distinction between the kind of language and associated behaviors that impart knowledge, the futile attempt at their codification *vis-à-vis* tautological formal languages independent of appeals to experience or intuition.

One might question where is the evidence in the first place that such assumptions exist within those concerned with modeling knowledge. Unfortunately my reply would be that examples abound.

Engle in a recent article demonstrates such a confusion, regarding the strength and aims of knowledge modelers (amongst other like minded individuals) to be the,

> *"modeling of the complete subject-knowledge of business users and departments at all corporate levels" (Engle, 2003)*

I take it by use of the term "modeling" Engle refers to the formal process of construction of a complete and systematic methodology for representing the "*complete subject-knowledge*". If not, how might we meaningfully talking about the different *kinds* of "subject-knowledge" in regard to the model. The complications generated by this aspiration are unequivocal.

Tom Finneran imagines the following,

> *"Knowledge is collected from all existing sources including people, systems, data stores, file cabinets and desktops. All knowledge of value is stored in the organizational knowledge repository. For virtual teams, this knowledge would be immediately conveyed to those*

people and systems that could use it. The right knowledge will go to the right person or system at the right time. Current knowledge can be retrieved from the system at any time in the future." (Finneran, 2000)

The conflation between the scientific principles that under-pin *data management*, which can be appropriately interpreted as information, captured within the confines of a logically complete formalized structure (namely a relational database) and the nature of knowledge is clearly evident here.

Barclay and Murray, knowledge management consultants add further fuel to the fire.

"knowledge management often encompasses identifying and mapping intellectual assets within the organization, generating new knowledge for competitive advantage within the organization, making vast amounts of corporate information accessible, sharing of best practices, and technology that enables all of the above — including groupware and intranets." (Barclay *et al*, 1997)

Conceptions of and attempts to model knowledge within current computational frameworks fraught with their requirements for formal codification and the associated applied formalisms seem doomed. If language cannot lend itself to determinism and while knowledge is language dependant then knowledge is in the mind of the beholder.

In summary;

- Data and thus the associated inferred appropriately interpreted information can be logically represented within formal systems (such as relational databases).
- Data represented in such a way involves the elimination of appeals to intuition and experience in its interpretation. Its computational representation is within a logically closed system.
- Knowledge inherently involves appeals to intuition and experience in order to confirm or detract in its confirmation.
- Knowledge is reliant upon natural language for its expression.
- Natural language in light of Quines indeterminacy arguments cannot be systematically and logically represented.
- Thus, knowledge cannot be logically and systematically represented.

The upshot of this analysis is we ought to abandon our typical conceptions of knowledge management which rely upon existing data and information oriented architectures, attempts to model knowledge within such systems is futile. Alternatively we drop the hype and be more honest about what we are presently doing which simply is another form of information management.

Bibliography

Alston, W., 1986, Quine on Meaning., in The Philosophy of W.V. Quine, exp edition, Hahn and Schlipp (ed.), Open Court.

Barclay, R.O., Murray, P.C., 1997, What is Knowledge Management.
http://www.media-access.com/whatis.html

Cohen, R.S., 1963, Dialectical Materialism, in The Philosophy of Rudolf Carnap, Schlipp, P.A. (ed.), Open Court, London.

Date, C., 2000, An Introduction to Database Systems, Addison Wesley Longman Inc.

Dummett, M., 1974, The Significance of Quine's Indeterminacy Thesis, Synthese 27, 1974, pp. 351-97. Reprinted in Truth and Other Enigmas, Duckworth, London, 1978 pp. 375-419.

Engle, P., 2003, Data Modeling, Left and Right., The Data Administration Newsletter (TDAN.com) http://www.tdan.com/i024hy03.htm

Gupta.A., McDaniel, J., 2002, Creating Competitive Advantage By Effectively Managing Knowledge: A Framework For Knowledge Management, Journal of Knowledge Management Practice, Volume 3, 2002. http://www.tlainc.com/articl39.htm

Finneran, T., 2000, A Component Based Knowledge Management System., The Data Administration Newsletter (TDAN.com) http://www.tdan.com/i009hy04.htm

Quine, W.V.O.,1960, Word and Object., The MIT Press, Massachusetts.

Reid, T., 1764, An Inquiry into the Human Mind on the Principles of Common Sense., Brookes, D.R. (ed.), Edinburgh University Press, 2000.

Waismann F, 1945, Verifiability, Proceedings of the Aristotelian Society, supplementary volume 19. reprinted in G.H.R. Parkinson (ed.), The Theory of Meaning (1968)

Wittgenstein, L. 1953. Philosophical investigations. (Anscombe, G.E.M., trans.). Basil Blackwell, Oxford.

(Footnotes)

[1] A variation of this idea was published recently in *The Journal of Knowledge Management Practice*

[2] *Normative statements* are "ought to" statements, in a business environment these statements usually lead to or are the outcome of strategic decision making. I use the term normative statements and value statements interchangeably.

[3] Knowledge usually adopts a more lofty stance than belief since we typically consider knowledge to be true, not ever false, whereas belief can be false. Knowledge in this sense is "justified belief". The method of justification for any knowledge is another matter – one that we won't discuss here.

[4] The reason for this relates to explanation. The explanation for a theoretical assertion *X* relates to the empirical evidence that supports *X* – this is the basis by which science proceeds.