Kim H. Veltman

"Challenges for the Semantic Web and Information Systems from Culture":

Abstract

This paper notes that 20th century advances in computer science and information systems narrowed the earlier scope of semantics and outlines six challenges for information systems to:
1) return to the earlier scope of semantics at the turn of the 20th century
2) address all six basic questions.
3) create cognitive models that link with the physical world
4) create entities to allow integration of changing paradigm shifts and world views
5) include multi-lingual and multi-cultural dimensions
6) develop models that reflect differences and limitations in media types.

At the turn of the 20th century, Semantics was a "science that studies the content (meaning) side of linguistic signs." It was linked with five other fields: 1) Semiotics a "science of general properties of sign systems"; 2) Lexicology, a "science that studies vocabulary of language also called Lexical Semantics"; 3) Lexicography, a "science of dictionaries and their creation"; 4) Semasiology, a "branch of semantics that seeks meaning departing from expression side of language that studies the development and changing meaning of words" and 5) Onomasiology, a "branch of semantics which departs from a meaning side of language and asks what expressions exist in other languages."[1]

During the 20th century there were many breakthroughs in computers and information systems. Paradoxically this quest to achieve automated forms of communication also led to a great simplification of semantic problems and a serious dumbing-down with respect to solutions.[2] As a first challenge, this paper argues that we need to integrate what we left behind if we are to go ahead. In so doing, information systems can meet a second challenge of going beyond Who? and What? questions to deal also Where?, When? How? and Why? questions.

In the 1940s, Claude Shannon, who worked with Vannevar Bush on the ENIAC (Electronic Numerical Integrator and Computer), chose Boolean logic[3] as his point of departure. This reduced logic to choices between three logical operators: *and*, *or* and *not*. In terms of the semantic primitives,[4] these operators dealt effectively with Exclusion, simplistically with Inclusion, and in a very limited way with Intersection. In terms of basic questions this focused on Who? and What? questions, ignoring entirely Where?, When?, How? and Why? questions.

In the 1950s, Curry and Feys developed typed combinatory logic.[5] This introduced a distinction between three levels of language: 1) Phenotype, which describes natural language as order of words; 2) Genotype, which expresses grammatical invariants and structures and 3) the Cognitive level which deals with lexical predicates as represented by semantic cognitive schemes. This linked cognitive science with semantics[6] and separated clearly perceptual and conceptual dimensions.[7] It also focussed so much on mental dimensions that a recent survey of trends could ask: Does representation need Reality?[8] Hence, a third challenge is to remember that our models need to explain the physical world and to bridge physical and mental worlds.[9]

In the 1980s, the influential work of Langacker established the foundations of a new cognitive grammar.[10] This introduced a fundamental distinction between nominal predications used to designate things (i.e. nouns) and relational predications used for states (a-temporal relations) and processes (adjectives, adverbs, prepositions and verbs). In terms of basic questions, this established a distinction between Who? and What? (nouns) and relations entailing Where?, When?, How? and Why? (adjectives, adverbs, prepositions and verbs). This prompted new attention to relations using verbs: troponymy.[11] It also obscured very clear relations that had been established between division and partition. Since the 1980s, traditional distinctions have been even further eroded. For instance, Eduard Hovy, [12]one of the authors of the influential, *WordNet* recently noted (2002):

> We define an ontology rather loosely as a set of terms, associated with definitions in natural language (say English) and, if possible, using formal relations and constraints, about some domain of interest, used in their work by humans, data bases and computer programs. We view a set of semantic relations organized into collections and perhaps related in a generalization hierarchy as a special instance of an ontology.[13]

These definitions are important because they establish new links between ontology and semantics that help to account for the meteoric rise of these two concepts in the past decade. Meanwhile, Christiane Felbaum (2002) claimed that: "If one examines the lexicalized concepts in relation to one another, it becomes clear that they differ in systematic ways that are characterizable in terms of similarities and contrasts. These consistent differentiations among concepts are what we call *semantic relations*."[14] Interpreted narrowly this definition would mean that semantics has nothing to do with who or what a thing is, or what it means. Rather, semantic relations would be limited to differentiations in terms of similarities and contrasts.[15]

To understand the larger challenges a brief excursus in history is necessary. In Antiquity the study of language was divided into the structure of language (grammar); the logic of language (dialectics, later logic) and the effects of language (rhetoric). This *trivium* of grammar, dialectic and rhetoric formed the language side of the seven liberal arts.[16] What? questions were shared by Grammar and Dialectic (Logic, figure 1). How? questions were the domain of rhetoric. Why? questions were relegated to philosophy. This discipline based approach meant that it was not until the 19th century that linguistic

Discipline	**Grammar, Dialectic**	Geography	History	**Rhetoric**	Philosophy
Function	Structure, Logic			Effects	Causes
Question	Who?, What?	Where?	When?	How?	Why?

Figure 1. Basic questions and basic disciplines.

variation and dialect studies (Where? questions) and etymology (When? questions) came into focus. The 20th century precursors to the semantic web returned the focus to logic (What? questions). As a result, semantics as a study of the meaning of words, language and sign systems narrowed to become a study of relations and relationships between static words and terms. Hence we need to expand the scope of semantics to cover all the basic questions.

The focus on logic has had a more subtle consequence of assuming that the semantic web is/should be about a single logic, namely, the truth system with which we happen to use today. In the realms of science and business this is completely 'logical'. When making transactions, we need assurances that our customers use the same rules as we do.

Inherent in this approach is the assumption that an entity is substantially one thing with fixed properties: e.g. the sun is a body around which planets rotate as the heliocentric theory claims. If we decided that today's heliocentric explanation were the sole admissible theory, then we would theoretically be unable to understand and to communicate the views of all proponents of a geocentric universe prior to Copernicus in the West and even more recently in some isolated regions of the world. Hence a fourth, more elusive challenge is integrate different paradigm shifts and world-views into our plans for the semantic web and information systems by developing concepts of changing entities. We need to develop the plural meaning of information *systems*.

Related to this is a problem that the official national view is often quite different from a regional or local view. The explanations of Indian New Delhi concerning Kashmir are typically different than the explanations of Pakistan's Karachi or Kashmir's explanations of itself. Their language, claims, and even their maps may be different. This leads to a fifth challenge whereby our semantic webs and information systems need to become more multi-lingual[17] and multi-cultural.[18]

One of the important insights of the late Marshall McLuhan was that every medium for communicating knowledge also shapes and limits the knowledge that it communicates. McLuhan argued that new media typically used the previous medium as their content. Hence, the telegraph used print as its content, print used manuscripts as its content; manuscripts used oral communication as their content and so on. This was one of the reasons for McLuhan's claim that: "the medium is the message."

Digital communication is not simply another step in the tradition of a new medium adopting the content of its predecessors. Once materials are digital they can be translated back to other media without difficulty. Potentially, a digital text can be output as a printed text, as oral-audio text or, using stereo-lithography, even as cuneiform text.[19]

Implicit in this insight is a sixth challenge: How can we create a more comprehensive semantic web with information systems that both reveal these new potentials and allow us to recognize explicitly the differences (both qua strengths and limitations) of earlier media?

The nexus of ontologies and the semantic web mentioned earlier led to a focus on metadata. The scope of metadata needs to be expanded into virtual reference rooms (which link classification systems, dictionaries, encyclopaedias and titles)[20] that provide new access to distributed digital libraries and fora for collaborative research and creativity: a new virtual agora. These can evolve into a Distributed European Electronic Resource (DEER) [21] and lead to global efforts in the form of a World Distributed Electronic Resource (WONDER).

Acknowledgements

Notes

[1] See: http://www.uni-bonn.de/~dbuncic/14ung/semantik.pdf

[2] The ideas in this paper are discussed in more detail in the author's "Towards a Semantic Web for Culture," *Journal for Digital Information,* Summer 2004 (in press).

[3] Boolean logic diagrams are themselves a simplification of Euler's diagrams. For Euler and Quantified Expressions
See: www.rci.rutgers.edu/.../Deduction/ EulerDiags.html

[4] John Sowa, "Ontology, Metadata, and Semiotics": Presented at ICCS'2000 in Darmstadt, Germany, on August 14, 2000. Published in B. Ganter & G. W. Mineau, eds., *Conceptual Structures: Logical, Linguistic, and Computational Issues*, Lecture Notes in AI #1867, Springer-Verlag, Berlin, 2000, pp. 55-81.John F. Sowa, Ontology Metadata and Semiotics.
See: http://users.bestweb.net/~sowa/peirce/ontometa.htm.
For a more detailed understanding of his work see: John F. Sowa, *Knowledge Representation: Logical, Philosophical, and Computational Foundations*, Brooks/Cole Publishing Co., Pacific Grove, CA, 2000.
For a fuller bibliography cf. John F. Sowa.
See: http://www.jfsowa.com/pubs/index.htm
As noted by Sowa, these semantic primitives go back to Peirce in his semiotics.
See: http://users.bestweb.net/~sowa/peirce/ontometa.htm. Peirce claims that there are five semantic primitives: Existence, Coreference, Conjunction, Negation and Relation. As will become clear in this essay, this overlooks that the other semantic primitives are also formal relations.

[5] H. Curry and R. Feys, *Combinatory Logic*, Amsterdam: North Holland, 1958.

[6] For a thoughtful essay on this subject cf. Mihailo Antović, "The Position of Semantics within Contemporary Cognitive Science," UDC 81'37:165.19: in: *Facta Universitatis, Series Linguistics and Literature*, vol. 2, no. 10, pp. 415-424.

See: http://facta.junis.ni.ac.yu/facta/lal/lal2003/lal2003-06.pdf
For linguistics and cognitive science
   See: http://www.press.uchicago.edu/Subjects/virtual_linguistics.html
[7] S. Pribbenow, "Meronymic relationships", in *Semantics* as in note 8, p. 39.
[8] Conference in Vienna May 14-16 1997.
   See: http://www.univie.ac.at/cognition/conf/ntcs97/
[9] For an historical discussion see: "Visualization and Perspective. Visualizzazione e prospettiva" *Leonardo e l'eta della ragione*, eds. Enrico Bellone e Paolo Rossi, (*Scientia*, Milan, 1982), pp.185-210 (English), pp. 211-224 (Italian).
[10] R. W. Langacker, *Foundations of Cognitive Grammar*, Stanford: Stanford University Press, 1987, (Volume 1: *Theoretical Prerequisites*) .
[11] Cf. Christine Felbaum, "On the Semantics of Troponymy, in: *The Semantics of Relationships*, ed. Rebecca Green, Carol A. Bean, Sung Hyon Myaeng, Dordrecht: Kluwer Academic Publishers, 2002, pp. 23-34.
[12] For Further publications
   See: http://www.isi.edu/natural-language/people/hovy/publications.html
[13] Eduard Hovy "Comparing Sets of Semantic Relations in Ontologies," in: *Semantics,* as in note 8, 2002, p.92.
[14] Christiane Felbaum, "as in note 8, p.23.
[15] To be sure there are exceptions to this trend. Cf. Christophe Jouis, "Logic of Relationships," in: *Semantics*, as in note 8, 2002, especially pp. 129-130.
[16] Along with the *quadrivium* of mathematics, arithmetic, music and astronomy.
[17] Cf the important work of *Accès Multilingue au Patrimoine* (AMP). AMP newsletter under Ministère de la Culture, France.
   See: http://www.culture.gouv.fr/culture/mrt/numerisation/fr/f_01.htm.
   On this subject see also: Marc van Campenhoudt, *Abrégé de terminologie Multilingue*,
   See: http://www.termisti.refer.org/theoweb1.htm#intro
[18] From a larger historical viewpoint there was a brief time in the 19th and early 20th centuries when simply exporting one's own political solutions abroad via colonialism and imperialism seemed a viable solution to differences. Today, however, when Europe is less than 5% of the world population, how can we legitimately assume that our version of politics is necessarily the only model when countries such as China, now the 8th largest economy in the world, have other systems? This is in no way to claim that we should abandon democracy, but simply to urge for multiple logics, belief systems and models.
[19] Cf. the author's "Electronic Media and Visual Knowledge," *Knowledge Organisation*, Wurzburg, Vol. 20 (1993), No. 1, pp. 47-54.
[20] See: http://www.i-massweb.org/
[21] Cf. the study by Francesca Monti and Suzanne Keene on the DEER under E-Culture Net. See: http://www.eculturenet.org/FP5/publicPDF/deliverable11a.pdf.