SOFTWARE ARCHITECTURE WITH EMERGENT SEMANTICS
How can systems be weakly coupled, but strongly referenced

Len Yabloko
Next Generation Software

Keywords: Semantic Web, Software Architecture, Grid Computing.

Abstract: This paper offers a unified modelling and computing paradigm with explicit (i.e. traceable to the facts) epistemological support of semantic primitives. It builds on the method of logic stratification and alignment first proposed by the author that allows to overlay the global computational grid with meta-space based on distributed shared content addressable memory (a.k.a. Tuple Space). This paper shows a strong connection between semantic, algebraic and topological properties of such meta-space, which makes it ideal conduit directly relating semantics of primitives to grounding of enterprise processes in the global computational grid.

1 INTRODUCTION

Emergent Semantics is a form of interoperability in open and dynamic environments, such as global computing grid, which is achieved incrementally with negotiations conducted to reach agreements on common interpretations within the context of a given task (Aberer, 2004). Following is a brief outline of its principles:
1) Agreements fulfill a protocol upon which shared dynamic context can be constructed;
2) Assumptions must be constantly validated;
3) Global agreements result from aggregations of local agreements;
4) Local interactions occur in the global context;
5) Agreements induce semantic self-organization;
Informally, self-organization can be characterized by a complete distribution of control and by the restriction to local interactions, information and decisions. Global structures can then emerge from such local interactions.
One particularly successful example of emergent semantics is link-based ranking as used in Google. A global semantic agreement is obtained on “general importance” of Web documents. We can compare Google’s approach to the one taken by Web directories, such as Yahoo. In web directories the decision on importance of Web documents is taken globally, manually and centrally. This clearly limits the scalability of the approach.
A next natural step beyond ranking-based methods, which ignore the structure of the content, would be to apply the principle of emergent semantics to obtain interpretations for structured data. One possible avenue of how this might be achieved is currently being opened in the area of peer-to-peer (P2P) data management, where local schema mappings are introduced in order to enable semantic interoperability (Majkic, 2004). Such local schema mappings perform as local communication mechanisms for establishing consensus on the interpretation of data. Once such infrastructures are in place, the principles of emergent semantics become directly applicable, and semantic structures can build up from large numbers of purely local, pair-wise interactions.
For practical reasons it is not possible to include in a database all the facts pertaining to the objects in a given application domain. The Closed World Assumption (CWA) simply declares that all relevant facts are stored in the database, so that any statement that is true about the actual world can be deduced from facts in the system. CWA in computer and physical science have a lot in common, not only because both are simplifying assumptions, but also because in both cases the answers to some of the most fundamental problems are deferred for pragmatic reasons. In computers science such problem is known as “Frame Problem”: under the CWA any dynamic reasoning becomes non-monotonic, requiring revision of a database (or more broadly a belief base) for every new observation. Simply not doing so is often (in error) considered a monotonic reasoning. And this is why conventional

Yabloko L. (2005).
SOFTWARE ARCHITECTURE WITH EMERGENT SEMANTICS - How can systems be weakly coupled, but strongly referenced.
In *Proceedings of the Seventh International Conference on Enterprise Information Systems*, pages 283-288
DOI: 10.5220/0002529402830288

information technology runs into the Frame Problem when used on the Web. That fact resulted in current attempts to develop so called "Semantic Web", which allows correct monotonic reasoning over semantically fixed information, combined with non-monotonic reasoning required by Web's emergent semantics. Semantic Web must provide following:

– The ability to compute the consensus semantics or reality based on an analysis and aggregation of the individual events observed.

– The ability to establish and maintain the scope if each premise, which implies the ability to establish the (global) identity of its source.

The rest of this paper is organized as follows. Section 2 will focus on applying well known results of research in non-monotonic reasoning to Emergent Semantics. Section 3 will apply very recent results in non-classical logic and Category Theory to Emergent Semantics. Section 4 will introduce the software architecture proposed by the author, and Section 5 will discuss future possibilities.

2 LOGIC STRATIFICATION

In First-Order Logic (FOL), laws and facts are propositions, and there is no special mark that distinguishes a law from a fact. To distinguish them, a context mechanism is necessary to separate first-order reasoning with the propositions from meta-level reasoning about the propositions and about the distinctions between laws and facts. This kind of separation in logic is called *stratification*.

The author of present paper had proposed (Yabloko, 2003) a method and algorithm for logic stratification and alignment that is very much in line with (Sowa, 2002) proposal of stratifying non-monotonic reasoning into FOL reasoning over the beliefs represented in ontologies with graph-based meta-theory of contexts. It's major novelty, however, is in direct epistemological support for semantic primitives.

For illustration of its basic principle consider an open set of possible worlds $\boldsymbol{w}_{o,}$ $\boldsymbol{w}_1$ $\boldsymbol{w}_2$ ... with accessibility relation **R** and semantic primitives ***a,b***, …, which form the alphabet shared among different worlds. Although the meaning of each symbol is grounded differently in each world, that is the symbol can be used for very different purposes and have very different significance associated with it, all symbols relate to each other in a very specific way in each world. Now, let's assume that the only requirement for a "safe passage" from one world to another is that relations among some symbols (of our choice) do not contradict those in other worlds. This situation is not very different from the one that take place when boarding an airplane: each passenger carries his items of significance by the screening machine at the gate, which determines if those items are safe for flying. If certain item is marked unsafe, then the passenger has a choice to leave it at the gate or perhaps to travel by car. The transport mechanism between the worlds, proposed by the author, is called "causal stream" because the relations in question are causal relations between the events represented by each symbol. Note that these relations do not have to be direct. The requirement is that of all possible event chains in each world - none violates causal relations between the given subset of events. This requirement easily translates into mathematical notion of partial order. In fact, the entire idea is a variant of more general idea, developed in Domain Theory (Karazeris, 2001), of representing a process as a characteristic function measuring the extent to which the event (computation path) shapes can be realized by the process.

The idea of "causal stream" has additional epistemological significance and can be used for consensus derivation in the emergent semantics scenario. There it provides a direct connection between semantic grounding of a process and its physical (e.g. computational) properties. So negotiating agents can adequately access the consequences of their decisions, and not merely its semantics. Dually, it reinforces the semantic primitives with epistemological support in a form of concrete physical properties of the word, which they can express. That epistemological link defines the architectural term "strong reference", coined by the author of present paper to emphasise its effect on systems architecture, which becomes less brittle as a result of adding plasticity of emergent semantics. It is important to note, however, that the fundamental idea of linking semantics of an expression with its pragmatics (ie. usefulness) was introduced by Frege in late 19^{th} century as a sense of the expression or its "mode of presentation".

The above example shows how the belief revision in non-monotonic reasoning can deal with uncertainty resulting from incomplete knowledge. Indeed, the transition from one possible world to another did not require a complete knowledge revision, but only affected that knowledge relying on a limited subset of semantic terms. Moreover, the scope of such revision is easily computable for as long as the knowledge base is terminologically indexed, which is a common technique, known as faceted taxonomy. Thus, agents in emergent semantics scenario can perform tractable reasoning and focused belief revision under CWA.

3 SEMANTIC AGGREGATION

Essentially, aggregation is a form of meta-level reasoning. In computer science meta-level reasoning was introduced by the Type Theory, which is based on the earliest form of logic stratification - called ramification. The notion of logical context, or rather of its philosophical predecessor – *sense*, introduced by Frege, is totally absent in Russell's theory. It is replaced by the notion of the logical order of proposition, and that has very profound implications for semantics. Tarski's Model Theory with its declarative semantics can be seen as an attempt to re-introduce the real world context into the propositional logic.
It is worth recalling that classical (propositional) logic came out of original "naïve" set theory as a solution to a logical paradox in Frege's "Foundations of Arithmetic", which Russell had discovered and solved by means of ramification. The paradox, however, was not exactly solved, but rather deferred. In essence, the paradox, which Russell himself called "vicious-circle principle," that is, "no totality can contain members defined in terms of itself", is a basic paradox of self-reference. Russell's solution transforms the totality into a logical equivalent of mirrors hanging on both sides of a long corridor (infinitely long in general case): each mirror showing the picture of another mirror along and across the corridor (such system of mirrors was invented by Egyptians to bring the light deep inside the famous piramids.) If each mirror is a set of propositions, then its logical order corresponds to the order of the mirror relative to the source of light. The question deferred by this solution is: how to transform logical propositions (corresponding to observed light) into a theory (corresponding to the World outside.) No such transformation is known!
But if we take what is arguably the most important application of Set Theory: Relational Algebra, and look at its most basic construct: binary relation from the functional point of view provided by Category Theory, we will discover a natural transformation called *pullback*, which can be applied inductively to transform any set of binary relations into a theory. Scope does not allow us to provide a detailed explanation of that fact, but a theory constructed this way is intentional in that it applies to the extension (i.e. the World) up to *isomorphism* (Vigna, 2002). The inverse is also true: when we obtain the extensional knowledge by making and observation, the intentional a-priory cause is not knowable because the observation itself is its causal component. Pullback transformation corresponds to irreversible process of semantic aggregation.

Aristotle in "The Categories" includes what we see as effects (the results or consequences of the event) as causes. His metaphysical framework included four types of cause or explanation:
• ***E*** *efficient - that which makes a change happen,*
• ***M*** *material - what the change happens to,*
• ***F*** *formal - what the change results in, and*
• ***N*** *final - the end or purpose of the change.*
Figure 1 represents a categorical pullback of binary multi-relations corresponding to four-way causality:

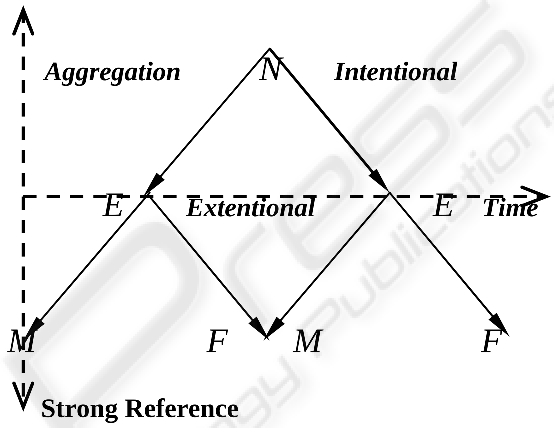

Figure 1: Causal stream

Conjecture 1*: For any process, represented as causal strem, there is a group of natural transformations which does not violate the stream. This group is called " strong reference".*

4 SOFTWARE ARCHITECTURE

Global Grid protocol architecture has five layers: fabric, connectivity, resource, collective and application (Foster, 2001). Only "collective" layer includes some level of coordination between services and resources. This situation is sometimes called "hour-glass" architecture, in which the impact of application and fabric on each other is not controlled. This is the root cause of brittleness.
This paper proposes an alternative architecture that removes the "most fragile" point of "hourglass" by reinforcing collective layer with external "feedback" loop (Fig. 2) corresponding to the strong reference. Such feedback guaranties accurate cross-reference from different domains in a virtual organization.

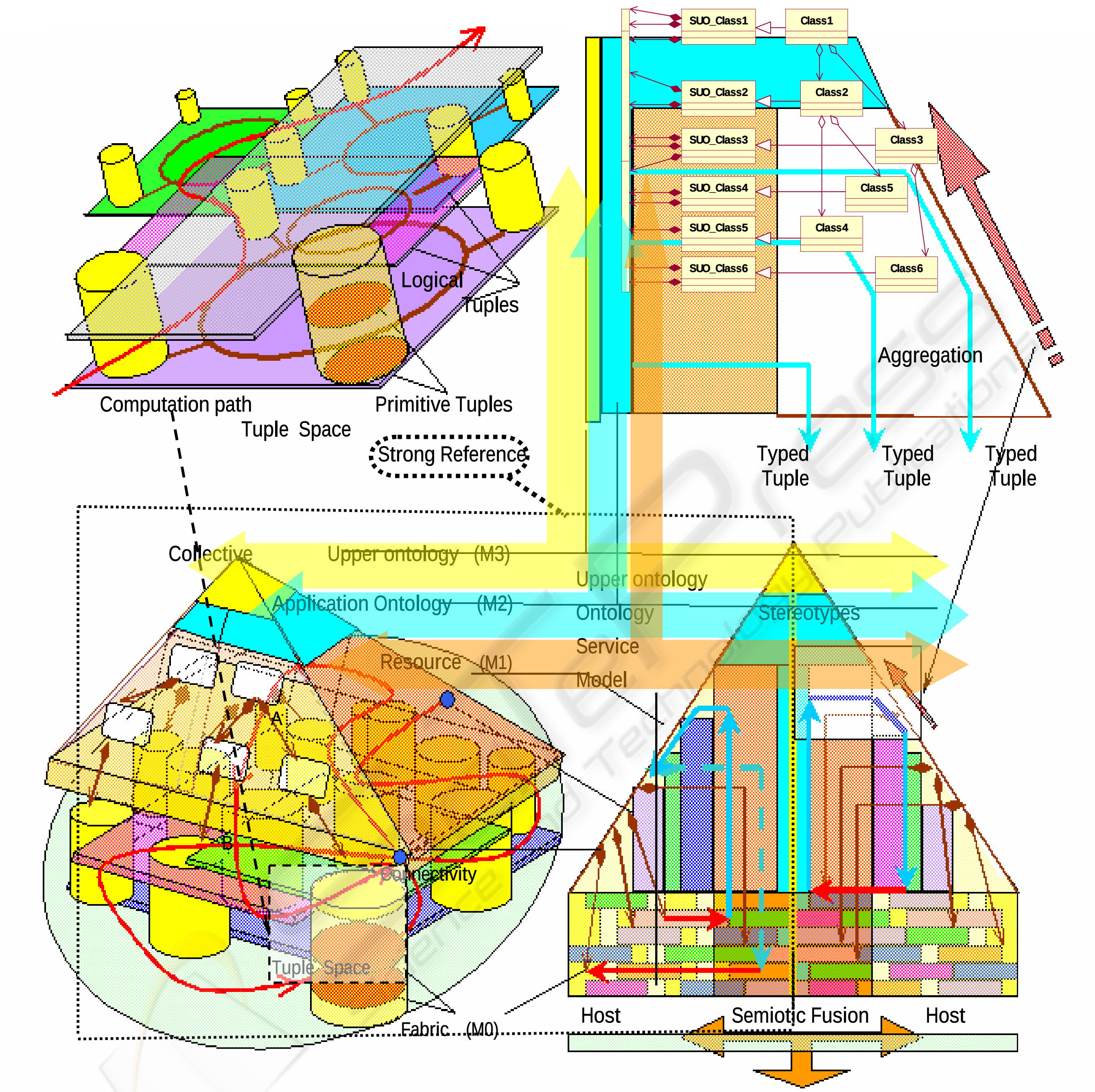

Figure 2: Software architecture with emergent semantics. Bottom left (a): Logical organization of the Grid using M0-M1 stratification (4L-MDA). Bottom right (b): Software stratification within a single domain (ie. single side of the Grid Pyramid). Top left (c): Tuple Space serving as strong reference between the fabric and collective layer of the Grid. Top right (d): Logical grounding of the resource and application layers of the Grid in federated Tuple Space with typed tuples born by the connectivity layer of the Grid

It serves as reinforcing "fibre" that reduces system brittleness. Classic four-side pyramid is shown only for simplicity. In fact the number of domain is not limited. Domains are separated by vertical edges of pyramid that correspond to accessibility relations between domains. The cylinders supporting the pyramid illustrate connectivity to the fabric.

4.1 Service Grounding

Grounding connects the service model to communication-level protocols, such as the message description of Web Services Description Language (WSDL). Although, the binding of atomic process to WSDL operation is limited by XML-based communication protocol (such as SOAP), this binding does not preclude any additional bindings of service to context (Ankolekar, 2001).

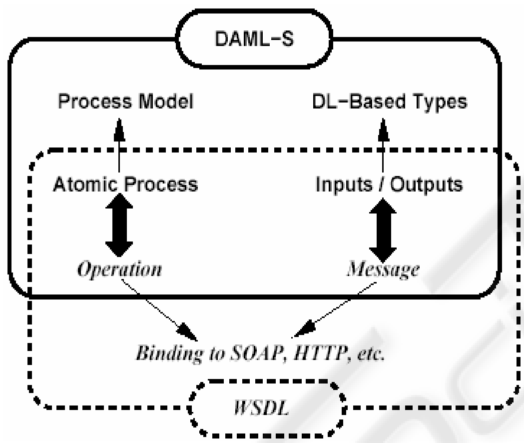

Figure 3: DAML Service grounding

Onto-Space provides additional binding of operation to the context, and then uses Tuple Space coordination mechanisms to enable strong reference.

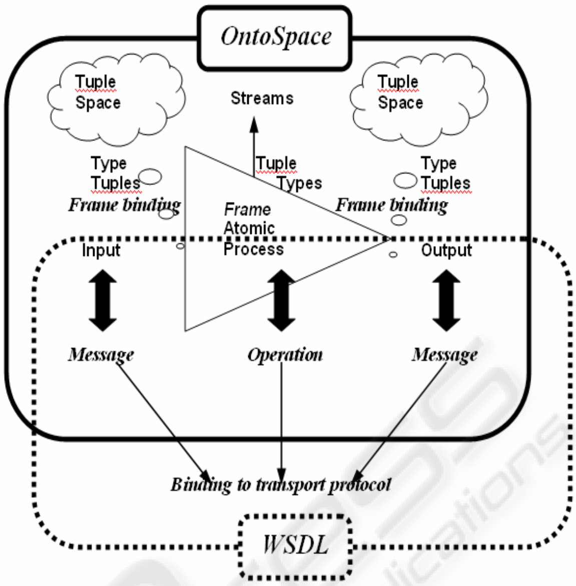

Figure 4: OntoSpace Service grounding

The binding of operation to the context is provided by type containers implemented by connectivity layer. The same or different communication protocol can be used, but the type systems used for this purpose must provide an accessibility relation to the service model. Semantic Web technologies seem to be ideal fit for that purpose allowing DL-based and other powerful type system (Horrocks, 2003) while using XML for communication. In the OntoSpace connected tuple spaces result in "streams" formed by multirelation spans across the different domains.
For example, a frame-based ontology can support a type constraints and inference required to create simple Ontospace. Such system was, in fact, developed by the author to represent a relational database as a type container. It uses Protégé frame-based environment (Noy, 2001) to implement Tuple Space with logical tuples restricted to binary relations. In it both primitive and logical tuples are implemented as frames (Figure 11), Prolog interpreter is used to implement type constraints, and DL reasoner is used for type inference.

5 FUTURE WORK

It is an opinion of the author that future computational grid must serve as a medium for emergent semantics. As such it will need to implement its principles in universal and consistent manner. This paper introduced the idea of tuple type systems that can be build using frame-based containers (such as Protégé) in order to allow tractable semantic aggregation. The effective application of such semantic type containers will require standardisation of tools around the idea of strongly typed tuple space, similar to present standardisation efforts in software application (component) containers. Onto-Space project intends to bring together presently isolated standards from MDA and Semantic Web to that effect.

REFERENCES

Aberer, K., Catarci, T., Cudre-Mauroux, P., Dillon, T., Grimm, S., Hacid, M.S., Illarramendi, A., Jarrar, M., Kashyap, V., Mecella, M., Mena, E., Neuhold, E. J., Ouksel, A. M. , Risse, T., Scannapieco, M., Saltor, F., De Santis, L., Spaccapietra, S., Staab, S., Studer, R., and De Troyer O. (2004). Emergent Semantics. In *Proc. of the 1st International Conference on Semantics of a Networked World*; Lecture Notes in Computer Science;3226; pp.14--43; Springer; Heidelberg; ISBN: 3-540-23609-0.

Ankolekar, A., Hobbs, J. R., Burstein, M., Lassila, O., Martin, D., McDermott, D., McIlraith, S. A., Narayanan, S., Payne, T., Sycara, K.. (2001). DAML-S: Service Description for the Semantic Web, In *Proc.of ISWC 2002*, volume 2342 of *LNCS*, pages 348–363. Springer, 2002.

Foster, I., Kesselman, C., Tuecke, S. (2001). The Anatomy of the Grid: Enabling Scalable Virtual Organizations [Electronic version]. Intl. Journal of Supercomputer Applications, 15(3), 2001

Horrocks, I., Patel Schneider, P. F. (2003, May). Three Theses of Representation in the Semantic Web. In *Porc. of the 12th International World Wide Web Conference: WWW2003*

Karazeris P. (2001). Categorical Domain Theory: Scott Topology, Powercategories, Coherent Categories [Electronic version]. *Theory and Applications of Categories,* Vol. 9, No. 6, 2001, pp. 106–120.

Majkic, Z. (2004, August). Weakly-coupled ontology integration for P2P systems. In *Proc of 1st Workshop on Peer-to-Peer Knowledge Management: P2PKM'04*

Noy, N. F., Sintek, M., Decker, S., Crubezy, M., Fergerson, R. W. & Musen. M. A. (2001). Creating Semantic Web Contents with Protege-2000. IEEE Intelligent Systems 16(2):60-71.

Sowa, J. (2002, May). Laws, Facts, and Contexts: Foundations for Multimodal Reasoning. *Knowledge Contributors*, edited by V. F. Hendricks, K. F. Jørgensen, and S. A. Pedersen, Synthèse Library vol. 322, Kluwer Academic Publishers, Dordrecht, 2003. Retrieved from http://www.jfsowa.com/pubs/laws.htm

Vigna, S. (2002). Multirelational Semantics for Extended Entity-Relationship Schemata With Applications [Electronic version]. In *Proc. of the 21st International Conference on Conceptual Modeling - ER 2002* Lecture Notes in Computer Science;2503; pp.35--49; Springer; Heidelberg; Retrieved July, 2004 from http://vigna.dsi.unimi.it/ftp/papers/Multirelations.pdf