EMERGENT SEMANTICS

Philippe Cudré-Mauroux
MIT Computer Science and Artificial Intelligence Laboratory
Cambridge, MA – USA
pcm@csail.mit.edu

SYNONYMS

Bottom-Up Semantics; Evolutionary Semantics

DEFINITION

Emergent semantics refers to a set of principles and techniques analyzing the evolution of decentralized semantic structures in large scale distributed information systems. Emergent semantics approaches model the semantics of a distributed system as an ensemble of relationships between syntactic structures. They consider both the representation of semantics and the discovery of the proper interpretation of symbols as the result of a self-organizing process performed by distributed agents exchanging symbols and having utilities dependent on the proper interpretation of the symbols. This is a complex systems perspective on the problem of dealing with semantics.

HISTORICAL BACKGROUND

Syntax is classically considered as the study of the rules of symbol formation and manipulation [8]. Despite its wide usage in many contexts, the notion of *semantics* often lacks a precise definition. As a least common denominator, it can be characterized as a relationship or mapping established between a syntactic structure and some domain. The syntactic structure is a set of symbols that can be combined following syntactic rules. The possible domains that are related to the symbols may vary. In linguistics, such domains are typically considered as domains of conceptual interpretations. In mathematical logic, a semantic interpretation for a formal language is specified by defining mappings from the syntactic constructs of the language to an appropriate mathematical model. Denotational semantics applies this idea to programming languages. Natural language semantics classically concerns a triadic structure comprising a *symbol* (how some idea is expressed), an *idea* (what is abstracted from reality) and a *referent* (the particular object in reality) [9].

Emergent semantics expresses semantics through purely syntactic, recursive domains. The notions underlying emergent semantics are rooted in computational linguistics works relating semantics to the analysis of syntactic constructs. In his seminal work on syntactic semantics [10, 11], William J. Rapaport defines semantic understanding as the process of understanding one domain in terms of another – antecedently understood – domain. This further raises the question of how the antecedently domain is itself understood. In the same vein, the antecedently understood domain has to be understood in terms of yet another domain, and so on and so forth recursively. To avoid to ground the recursion to a hypothetical base domain, Rapaport suggest the notions of *semantics as correspondence*, and lets the semantic interpretation function recursively map symbols to themselves or to other symbols. By considering the union of the syntactic and semantic domains, Rapaport regards semantics as syntax, i.e., turns semantics into the study of relations within a single domain of symbols and their interrelations. A dictionary is a simple example of a construct based on that paradigm, where the interpretations of symbols (i.e., words) are given by means of the same symbols, creating a closed correspondence continuum. Emergent semantics applies the conception of a closed correspondence continuum to the analysis of semantics in distributed information systems, by promoting recursive analyses of syntactic constructs – such as schemas, ontologies or mappings – in order to capture semantics.

SCIENTIFIC FUNDAMENTALS

Beyond its implication in linguistics – where it is conjectured that human beings inevitably understand meaning in terms of syntactic domains – emergent semantics is considered as being mostly relevant to computer science. Programs, database schemas, models, or ontologies have no capacity (yet) to refer to reality. However, they have various mechanisms at their disposal for establishing relationships between internal symbols and external artifacts. In the settings where humans provide semantics, relationships among symbols – such as constraints in relational databases – are means to express semantics. In order to rectify some of the problems related to the implicit representation of semantics relying on human cognition, some have proposed the use of an explicit reference system for relating sets of symbols in a software system. Ontologies serve this purpose: an ontology vocabulary consists of formal, explicit but partial definitions of the intended meaning of a domain of discourse. In addition, formal constraints (e.g., on the mandatoriness or cardinality of relationships between concepts) are added to reduce the fuzziness of the informal definitions. Specific formal languages (such as OWL) allow to define complex notions and support inference capabilities. In that way, explicitly represented semantics of syntactic structures in an information system consist of relationships between those syntactic structures and some generally agreed-upon syntactic structure. Thus, the semantics is itself represented by a syntactic structure.

In a large scale distributed environment of information agents, such as in the Semantic Web or Peer-to-Peer systems, the aim is to have the agents interoperate irrespective of their initial vocabularies. To that aim, an agent has to map its vocabulary (carrying the meaning as initially defined in its *base* schema or ontology) to the vocabulary of other agents with which it wants to interoperate. Hence, a relationship between local and distant symbols is established. This relationship may be considered as another form of semantics, independent of the initial semantics of the symbols. Assuming that autonomous agents acquire vocabulary terms through relationships to other agents and that agents interact without human intervention, the original *human assigned* semantics would loose its relevance; from an agent's perspective, *new* semantics would then result from the relationships to its environment. This is a novel way of providing semantics to symbols of agents relative to the symbols of other agents with which they interact. Typically, this type of semantic representation is distributed such that no agent holds a complete representation of a generally agreed-upon semantics.

From a global perspective, considering a society of autonomous agents as one system, one can observe that the agents form a complex, self-referential and dynamic system. It is well-accepted that such systems often result in global states, which cannot be properly characterized at the level of local components. This phenomenon is frequently characterized by the notion of *self-organization*. Thus, emergent semantics is not only a local phenomenon, where agents obtain interpretations locally through adaptive interactions with other agents, but also a global phenomenon, where global semantics emerge from the society of agents and represent the common, current *semantic agreement* in the system. This view of semantics as the emergence of a distributed structure from a set of dynamic processes – or more specifically as some equilibrium state of such processes – is in-line with the generally accepted definitions of emergence and emergent structures in the complex systems literature [2]. In that respect, emergent semantics can be related to dynamic systems disciplines such as evolutionary game theory, semiotic dynamics [13], or graph evolution.

KEY APPLICATIONS*

Multimedia Systems: Distributed multimedia applications were the first information systems to take advantage of emergent semantics principles in order to capture the semantics of shared objects. In this context, Santini *et al.* [12] argue that images do not have an intrinsic meaning, but that they are endowed with a meaning by placing them in the context of other images. From this observation, they propose a system where users are able to manipulate relations between images, and where the semantics of an image is emergent, in the sense that it is a product of the dual activities of users manipulating sets of images and of the database system. Along the same lines, Grosky *et al.* [5] focus on the role of context for giving meaning to a work of art. In their work, document semantics emerge through the analysis of users' browsing paths through a multimedia collection. They categorize the semantics of each item based on the collection of browsing paths.

Tagging Portals: Emergent semantics can be seen as a natural paradigm to analyze the distributed, user-driven process of giving semantics to items through the use of tags. Extreme Tagging [14] is a technique promoting the tagging of tags and the analysis of the relations between tags. Yeung *et. al.* [15] discuss how shared documents acquire meaning through their associations with other elements. In particular, they demonstrate how different meanings of ambiguous tags can be discovered through the analysis of a tripartite graph involving users, tags, and resources. Herschel *et al.* [6] discuss query expansion techniques by aggregating users opinions as tags and discuss the role of pragmatics in collaboratively creating semantics.

Heterogeneous Information Systems: Emergent semantics has been suggested as a way to capture semantics in decentralized and heterogeneous information systems [4]. In contrast to mediated integration architectures, recent decentralized integration architectures – such as Peer Data Management Systems – do not require the definition of any global schema or ontology. Thus, the global semantics of such systems can only be captured by considering the collection of conceptualizations as defined by the local databases, along with their interrelations. Semantic Gossiping [1] analyzes transitive closures of schema mappings in that context in order to infer semantic agreement. Related techniques suggest the use of graph theory or probabilistic networks [3] in order to capture global semantics in similar environments. More broadly speaking, emergent semantics techniques are increasingly being seen as a way to minimize manual input and maintenance when dealing with complex and heterogeneous information systems [7].

CROSS REFERENCE*

DATA INTEGRATION
MULTIMEDIA DATABASES
METADATA MANAGEMENT
PEER DATA MANAGEMENT SYSTEMS
PEER-TO-PEER DATA INTEGRATION
SCHEMA MAPPING

RECOMMENDED READING

[1] K. Aberer, P. Cudré-Mauroux, and M. Hauswirth. The Chatty Web: Emergent Semantics Through Gossiping. In *International World Wide Web Conference (WWW)*, pages 197–206, 2003.

[2] Y. Bar-Yam. *Dynamics of Complex Systems*. Perseus Books Group, 1997.

[3] P. Cudré-Mauroux. *Emergent Semantics: Rethinking Interoperability for Large-Scale Decentralized Information Systems*. PhD thesis, 3690, EPFL – Switzerland, 2006.

[4] P. Cudré-Mauroux *et al.* Viewpoints on Emergent Semantics. *Journal on Data Semantics*, VI:1–27, 2006.

[5] W.I. Grosky, D.V. Sreenath, and F. Fotouhi. Emergent Semantics and the Multimedia Semantic Web. *SIGMOD Record*, 31(4):54–58, 2002.

[6] S. Herschel, R. Heese, and J. Bleiholder. An architecture for emergent semantics. In *ER (Workshops)*, pages 425–434, 2006.

[7] B. Howe, K. Tanna, P. Turner, and D. Maier. Emergent semantics: Towards self-organizing scientific metadata. In *International Conference on Semantics of a Networked World (ICSNW)*, pages 177–198, 2004.

[8] C.W. Morris. Foundations of the theory of signs. *International Encyclopedia of Unified Science*, 1(2), 1938.

[9] C.K. Ogden and I.A. Richards. *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism*. Routledge & Kegan Paul Ltd., London, 10 edition, 1923.

[10] W.J. Rapaport. Syntactic Semantics: Foundations of Computational Natural-Language Understanding. In *Aspects of Artificial Intelligence (Dordrecht, Holland: Kluwer Academic Publishers)*, pages 81–131, 1988.

[11] W.J. Rapaport. What Did You Mean by That? Misunderstanding, Negotiation, ans Syntactic Semantics. *Minds and Machines*, 13(3):397–427, 2003.

[12] S. Santini, A. Gupta, and R. Jain. Emergent Semantics through Interaction in Image Databases. *IEEE Transaction on Knowledge and Data Engineering*, 13(3):337–351, 2001.

[13] L. Steels. Semiotic dynamics for embodied agents. *IEEE Intelligent Systems*, 21(3):32–38, 2006.

[14] V. Tanasescu and O. Streibel. Extreme tagging: Emergent semantics through the tagging of tags. In *International Workshop on Emergent Semantics and Ontology Evolution (ESOE)*, pages 84–94, 2007.

[15] C.A. Yeung, N. Gibbins, and N. Shadbolt. Understanding the semantics of ambiguous tags in folksonomies. In *International Workshop on Emergent Semantics and Ontology Evolution (ESOE)*, pages 108–121, 2007.