**Editor: Steffen Staab**
University of Karlsruhe
sst@aifb.uni-karlsruhe.de

# Emergent Semantics

John Naisbitt's saying, "We are drowning in information and starving for knowledge,"[1] is now a classic quotation for describing the Web, knowledge management, wearable computing, and e-learning—to name but a few. But, hey, isn't semantics en vogue? We're using semantic search tools that return precise answers, fast reasoning engines that exploit semantic relationships, and ontologies that integrate information. We are also investigating Semantic Web languages.[2] So, isn't the problem solved, at least in principle—give or take a little more programming effort?

Philosopher Ludwig Wittgenstein argued in *Tractatus Logico-Philosophicus* that language is composed of complex propositions that we can analyze into simpler ones until we arrive at elementary propositions.[3] He argued that the world is composed of complex facts that we can analyze into less complex ones until we arrive at atomic facts. The world is the totality of these atomic facts or "states of affairs," and elementary propositions logically stand for them.

Applying this logic to computer science, we could build intelligent systems that comprise elementary and complex propositions that stand for atomic and complex facts. Then, we could describe the semantics of the languages we use and employ query or reasoning engines to exploit our analysis. This resembles Wittgenstein's approach, which was to construct truth tables to analyze complex propositions.

However, Wittgenstein renounced much of his early work, saying his view of language (often referred to as Wittgenstein I) was too narrow and hence mistaken. In *Philosophical Investigations* (often referred to as Wittgenstein II), he argued that if we scrutinize how language is used, we might recognize its large variety.[4] Words can serve many different functions, thus we can use linguistic expressions in multiple ways. Although some propositions stand for facts, others are used to command, question, pray, thank, curse, and so on. This recognition of linguistic flexibility and variety led to Wittgenstein's concept of a language game—the sending and receiving of propositions—and his conclusion that people play different language games. The scientist, for example, is involved in a different language game than the theologian. The meaning of a proposition must now be understood in terms of its communication context—that is, it can only be described a posteriori. Eventually, the same might be said about the words people use to form propositions.

From this viewpoint, you might ask whether (most of) the semantics-based systems we have built exhibit some faults in their core principles. To some extent, this is not the case. In particular, semantic descriptions (such as ontologies) are models that provide abstractions of the world; naturally, they cannot reflect the wealth of all existing phenomena. This abstraction is fine if the systems run smoothly and do what they are intended to do. Rather than an insurmountable dichotomy between Wittgenstein I and Wittgenstein II, there might also be a synergy between the two. This is where emergent semantics might come into play.

Emergent computation is based on the idea that appropriate complex structures might arise purely from the physics of the task environment,[5] rather than from an architect's elaborate considerations. Such an evolution lets us build complex organisms without an explicit purpose of creating them.[6] However, emergent computation has typically been associated with evolutionary computation for optimization or combinatorics, which in turn is rarely applied to semantics.

Letting semantics emerge by observing interactions between human and machine agents, however, is an attractive and versatile paradigm for tackling several related problems:

- We can diminish the knowledge acquisition bottleneck that still hurts many of our semantics-based applications. Frequently, semantics-based systems don't run smoothly and suffer from a shortage of up-to-date semantic descriptions.
- We can observe many interactions between machine and human agents without additional human effort, letting semantic structures emerge on their own. Currently, these interactions often go unused.
- We can create a new basis for understanding natural language. For example, we could create a theory that better combines ideas from Wittgenstein I and Wittgenstein II.

Recent work, described in the following essays, has turned these high-level considerations into existing, running systems. Simone Santini explains how to use image retrieval from databases to exploit observations about system users and draw conclusions about the semantic content of images. Frank Nack's essay considers the more general area of multimedia, and Luc Steels discusses how to let semantics emerge between agents that communicate with made-up language.


1094-7167/02/$17.00 © 2002 IEEE

## Image Retrieval

Simone Santini, *University of California, San Diego*

Visual information retrieval has emerged in the last 10 years as a natural extension of certain database ideas to multimedia data—in particular, for images and video. The idea seemed natural in its simplicity: retrieve images from a large repository based on their content or, more precisely, certain standard interpretations of their contents.

Such a program's feasibility assumes that there is something we can reasonably call an image's *meaning*. Traditional computer vision hypothesized that we could, in principle, extract meaning from the image data and represent it in a symbolic or numerical way. Image retrieval makes the ontologically less committing hypothesis that syntactically measurable similarity would reveal similarity in meaning. In other words, it is possible to extract features from images and cast them into an appropriate metric space in such a way that similar images have similar meaning. The *query by example* model of image databases is based on this idea.

Although it makes a smaller ontological commitment, query by example still presupposes the existence of an image's meaning. Impossible as it might be to characterize this meaning using syntactic features, it is nevertheless still a function of the image data and, although absolute meaning can't be revealed, similarity of meaning between images can.

A fair share of the problems that plague image databases comes from this semantic presupposition, and we'll only solve these problems by redefining the concept and role of meaning in an information system.

### An image's meaning

I surmise that a fundamental difference between an image and a database record is their different status as signs. In particular, we can always view a database record as a preposition (that is, a *dicentic legisign*[1,2]), whereas images are terms predicated by an external discourse. The universe of images is contextually incomplete: taken in isolation, images have no assertive value but rely on some external context to predicate their content.

This characteristic is not exclusive of images: the same is true for normal speech and even fragments of formalized structures, such as databases. Consider the statement, "Louis XVI had to relentlessly fight for the success of the French revolution." The simplest question to ask about this statement is whether it is true or, in slightly more general terms, valid.

The answer depends. If the sentence appeared in a history book, it would certainly be false. Yet, in an English grammar book, as an example of a split infinitive, the sentence would be entirely appropriate. Although it doesn't refer to any historical fact, its validity would come from its being immersed in a different language game—that of valid grammar examples.

Of course, this is nothing new: the importance of context for interpretation has been recognized at various levels, from Umberto Eco's aphorism that "every decoding is another encoding" to the construction of the Cycorp knowledge base of "human common sense."[3] Eco's observation applies very well to the case of images, but the formalistic-reductionist approach at the basis of the Cyc attempt does not.

To show this, I present a formalized system in which meaning is strongly restricted, delimited, and formally described.

### Formalizing meaning

Consider the database table in Figure 1a. The records in Figure 1a don't mean anything, although most people will realize that the symbols in the first column are names, and Italian readers will probably realize that the symbols in the second column are fiscal codes. Apart from these intuitions, in a database, the meaning is formalized through a schema that, in this case, can be formatted as Figure 1b shows. With this schema, the information becomes a series of predicates such as "Dupont earns 95,000 US dollars per year, and her fiscal code is DPTDNS45E52B203K." The schema provides a framework for interpreting the records.

| | | |
|---|---|---|
| Berkowitz | BRKSML56D03D403K | 90000 |
| O'Malley | MLLRBR67M15F301A | 80000 |
| Dupont | DPTDNS45E52B203K | 95000 |

**(a)**

| | | |
|---|---|---|
| name:string | fiscal_code: fcodeType | salary:USD |

**(b)**

**Figure 1. (a) Information for a database; (b) an example in which the information's meaning has been formalized through a schema.**

Finally, Alexander Maedche elaborates on how the analysis of text might add to the semantic descriptions of the primitives in use. All discuss ways to let semantics emerge from simple observations from the bottom-up—rather than imposing concepts on the observations top-down—to provide precise query, retrieval, communication, or translation for a wide variety of applications.

Are you still drowning in information? Then, let's think about how semantics will emerge in your applications!

### References


1. J. Naisbitt, *Megatrends 2000*, Avon Books, 1991.
2. D. Fensel et al., "Trends and Controversies: The Semantic Web and its Languages," *IEEE Intelligent Systems*, vol. 15, no. 6, Nov./Dec. 2000, pp. 76–73.
3. L. Wittgenstein, *Tractatus Logico-Philosophicus*, Routledge, Cambridge, Mass., 2001.
4. L. Wittgenstein, *Philosophical Investigations*, Prentice-Hall, Upper Saddle River, N.J., 1999.
5. S. Forrest, ed., *Emergent Computation*, MIT Press, Cambridge, Mass., 1991.
6. R. Dawkins, *The Blind Watchmaker*, W.W. Norton & Company, New York, 1996.

What kind of meaning does this schema attach to the table? We can make many deductions based on these records. For instance, the rules governing Italian fiscal codes reveal that Dupont is a woman, born on 12 May 1945. Common conventions about naming babies make it clear that the names in the first column are with all probability the family name, and a general knowledge of heraldry suggests that Berkowitz's family is probably Jewish, O'Malley's is probably Irish, and Dupont's is probably French.

Are these observations included in the database semantics, and how is this decided? In a typical database, the operational aspects of meaning are determined by the data types that compose the various columns. A great deal of work at the conjunction between database theory and functional programming has been done to specify the structure of data types,[4] their semantics (through monoid or monad comprehension),[5] and their logic invariants (through the inclusion of theories in the specification of the data type).[6] In this way, the meaning of, say, the fiscal code is encapsulated in the algebra and theory that describe the data type fcodeType. These rules will state, for instance, that it doesn't make sense to add two fiscal codes, but it makes sense to compare the birth dates of two fiscal codes. Depending on the data type definition, it might not be possible to extract gender and birth date from a fiscal code.

The only semantics in the database are the algebraic and logic properties that the data types induce—all other interpretations are excluded. The foundational assumption of databases is that we can always determine this restriction in the data's meaning a priori, so that designing a schema, some data types, and some data algebras that completely encapsulate the desired semantics is possible.

I call this set of constructs, which determine the semantics of the records, the database's *discourse*. I use this term rather the more common *ontology* to emphasize that the semantic specification that I am considering has multiple aspects (logic, functional, algebraic) that the usual concept of ontology doesn't cover.

The previous observations amount to saying that in a traditional database, the database schema determines the discourse a priori, completely formalized, expressed in the same formalism in which the data are expressed. In language, on the other hand, the discourse is never completely formalized but can be encoded without leaving the system whose semantics we must determine. In other words, the contextual indicators of a sentence's semantics can be expressed in the same sign system in which the sentence is expressed (language).

## Discourse

What plays the role of discourse (in the sense indicated earlier) for images? Consider a picture of Umberto Eco conversing with St. Thomas Aquinas. Is this image valid? As in the sentence about Louis XVI, it depends. If the picture appeared on the front page of the *New York Times*, it would be considered a lie: Eco never had a conversation with Aquinas, and the front page of the *New York Times* is supposed to document real events.

On the other hand, had the same picture been published in the Book Review section of the *New York Times*, illustrating the cover of a book titled *Semiotics through the Ages*, the picture would have been entirely appropriate. The difference between these two situations is in the different set of cultural and social conventions that regulate the publication of pictures on a newspaper's first page and on its Book Review page. Similarly, a photograph placed in the context of a documentary is assumed to represent something real—not because of its contents, but because of the documentary photographer's social role.

So, we are once again in the situation in which some type of discourse or context is necessary for the interpretation. The difference between images and the discourses presented earlier is that, in this case, we can't express the discourse in the same terms as the data that it interprets.

## Image databases

Whereas the discourse of data typing in data, logic, or knowledge bases is always expressed in the same terms as the propositions whose validity it determines, this is not possible for image databases. In other words, a database schema is composed of terms of the same kind of the data that are in the database (strings, numbers, and so on) and can be manipulated using the same operations that manipulate the data. In an image database, these representations are radically different, because the context can't be represented as images.

Because of this, the program of visual information retrieval I mentioned in the introduction is, in its most ambitious form, unattainable. A pure repository of images, disconnected from any kind of external discourse, doesn't have any meaning that can be searched, unless we make some additional assumptions:

- The database is a priori inserted in a domain that is restricted enough to give meaning to a subset of the image content, and we can, for all practical purposes, disregard any other meanings of the images. This is the case for databases operating in certain domains such as medicine, where images are interesting only because of their diagnostic value. In this case, whether an MRI scan looks dramatic is irrelevant; only diagnostic features, such as those detecting a brain tumor, are of interest.
- The database is explicitly linked to an external discourse that can be formalized to a certain degree. This is the case of Web images: the Web's text and structure provide a reasonably rich environment in which we can identify topics and infer an image's intended message, at least to a certain degree.
- The user will endow images with meaning. The similarity relations that are valid between images in a given situation and that give meaning to images depend on the particular circumstances in which a given user or a group of users asks a given query. These circumstances are beyond the database's control and generally can't be formalized or encoded (since, as I argued before, the discourse about images can only be formalized outside of images but, ex hypothesis, this type of database contains only images). The meaning, in this case, does not reside in the database but is built by the user through appropriate associations during the image retrieval process.

These three assumptions result in three different search modalities—that is, in three different orientations of image database technology.[7] The third orientation is, in many respects, the most interesting. It is also the major contribution of image databases to the general field of information technology. In light of this, it seems reasonable that a semantic program for image databases should concentrate on the following points.

First, part of an image's semantics

derives from its relations with other images, which vary according to the query's particular circumstances. This implies that the semantics of images is at least in part functional and that a query process for image databases should manipulate similarity functions. An image database should include a complete algebra of similarity functions and treat similarity functions as first-class data.

Second, the semantics of the image's descriptors (features) should be specified, as much as possible, through a discourse (that is, through algebraic, logic, and functional means). However, this formalization will never be sufficient to delimit a semantics of interest—it will merely help in practical aspects of database organization[8] and support the user's true semantic-generating activity.

Finally, an image always has a meaning relative to the practices and social codes of a specific user. For example, two people in a picture can be judged too close (and therefore in a situation of intimacy) for an American viewer, but at a fair distance (and therefore in a situation of formality) for an Italian viewer, simply because the social code of spatial configurations is different in the two cases. In this sense, the goal of the interaction between the user and database is not so much to retrieve images based on a preexisting semantics but to create image semantics. The interaction itself is not configured as a query but as a navigation in which the user dictates similarities and associations between images and, through this activity, reorganizes the database to embody the desired semantic.

It is essential, for instance, that through the use of appropriate interfaces,[9] the user can decide which images are similar. This activity lets the database adapt its similarity measure to that which the user has in mind for that particular query. Consequently, the database can build, through repeated iterations, the semantics that the user has in mind for that particular query.

Relevance feedback has been a first step in this direction, but it is clear that to let alternative semantics emerge from the interaction between the user and database, the connection between the two must be much deeper. The user needs expressive means more powerful than simply selecting positive or negative examples, and the whole data organization inside the database should depend on the status of the interaction with the user.

The challenges that this organization will pose are at the boundary between database theory, image analysis, knowledge representation, and human–machine interaction. Developing solutions from such a maelstrom of different technical cultures and orientations will be an interesting and exciting experience.

## References


1. F. Merrel, *Semiosis in the Post-Modern Age*, Purdue Univ. Press, Lafayette, Ind., 1995.
2. U. Eco, *A Theory of Semiotics*, Indiana Univ. Press, Bloomington, Ind., 1976.
3. D. Lenant, *The Dimensions of Context-Space*, tech. report, Cycorp, 1998.
4. A. Albano, G. Ghelli, and R. Orsini, "Fibonacci: A Programming Language for Object Databases," *The VLDB J.*, vol. 4, no. 3, July 1995, pp. 403–444.
5. P. Buneman et al., "Principles of Programming with Complex Objects and Collection Types," *Theoretical Computer Science*, vol. 149, no. 1, Sept. 1995, pp. 3–48.
6. K. Didrich et al., "Programming in the Large: The Algebraic-Functional Language Opal 2," *Implementation of Functional Languages* (IFL'97), Lecture Notes in Computer Science, vol. 1467, Springer-Verlag, Berlin, 1998, pp. 323–338.
7. S. Santini, *Exploratory Image Databases: Content-Based Retrieval*, Academic Press, San Diego, Calif., 2001
8. S. Santini and A. Gupta, "An Algebra of Wavelet Features," *IEEE Int'l Conf. Multimedia and Expo* (ICME 2001), IEEE Press, Piscataway, N.J., 2001.
9. S. Santini, A. Gupta, and R. Jain, "Emergent Semantics through Interaction in Image Databases," *IEEE Trans. Knowledge and Data Eng.*, vol. 13, no. 3, May/June 2001, pp. 337–351.


## Media Information Spaces—A Semantic Challenge


Frank Nack, *CWI, Amsterdam*


The information society is leaving behind the cyberspace based on a hybrid system of traditional media (telephone, cinema, TV, theatre, museum, books, newspapers, and so forth) and digital information technology (networked and storage intensive computers, CD-ROMs, DVD, IP-telephony, Webcams, MIDI, and so forth). Rather, it is entering a knowledge space that facilitates new forms of creativity, knowledge exploration, and social relationships mediated through communication networks (including hypertext, interactive multimedia, interactive games, virtual reality, simulations, and augmented reality).

Such an interactive, open, and multimodal environment sustains the activation of the human and the artificial system's articulation powers to communicate ideas, where verbal, gestical, musical, iconic, graphic, or sculptural expressions form the basis of adaptive discourses. A basic aspect for such a space, which supports individuals but is still communal, is that information must be made accessible that is hidden in the unified structure of the single text, image, video, audio, or tactile unit. Thus, the goal is to create an environment in which media units and the relationships among them are understood as basic elements that can interrelate to produce new meanings.

To support this process of generating meaning, interpretation, and visualization, a system must know what is contained in the different media. For visual media, however, this poses a problem. Even though an image might provide a limited amount of visual information, it contains a wealth of meaning. This functionality is based on the two formal structures that can be assigned to every perceivable object in visuals: the *signifier* (which carries the meaning) and the *signified* (which is the concept or idea signified). The relation between the two elements is not a naming-process only, as the signified resembles not a thing but a concept. Secondly, the relation between the signifier and the signified is arbitrary. It is, in particular, the arbitrariness of the relationship between signifier and signified that enables the creation of higher-order sign systems and their diversity.

Thus, visual media requires more than characterizing its visual information on a perceptual level using objective measurements, such as those based on image or sound processing or pattern recognition. Creatively reusing material for individual purposes, which usually opens up questions of aesthetics and subjective interpretation, has a strong influence on the descriptions and annotations of visual media data, either created during the data's production process or added later. Providing semantic, episodic, and technical representation structures that can

change and grow over time is important. This also requires adaptable relations between the different type of structures.

## The Semantic Web

The Semantic Web is a first step toward addressing these problems (www.semanticweb.org). It should bring machine-processable content to Web pages, thus extending the current Web. The idea is to add ontology-based metadata to text or HTML documents to improve accessibility and provide a means for reasoning about the content. The applied technology is XML-based, which facilitates structural, cardinality, and datatyping constraints (XML Schema) on textual documents, allowing a comparison on structural levels. Richer semantic descriptions can be provided either as relation-oriented schemata (RDF, RDF Schema) or ontology-based technology (DAML+OIL). These technologies support in-depth indexing and classification of textual documents for presentation generation and navigation purposes.

To some extent, XML-based approaches also incorporate multimedia, either in the form of presentational languages such as Synchronized Multimedia Integration Language (SMIL) (integration of media style), SVG (with CSS for graphics), and XHTML (with CSS for formatted text), or transformational methods such as XSLT (document transformation) and CSS (control of style appearance).

However, the major drawback of XML-based environments is that they don't recognize visual media's dynamic nature or its variety of data representations and their mixes.

## MPEG frameworks

The Moving Pictures Expert Group is a working group of the International Organization for Standardization/International Electronics Commission. MPEG is in charge of developing standards for coded representation of digital audio and video, and it leads one of the broadest efforts in the direction of complex media content modeling. It aims to provide a framework for interoperable multimedia content-delivery services.

Semantic description languages have emerged in two of its standardization activities: in MPEG-4, as the Extensible MPEG-4 Textual Format (XMT) and in MPEG-7, as the Description Definition Language (DDL)—the multimedia content description interface.

In MPEG-4, the standard for multimedia on the Web, XMT provides content authors with a textual syntax for the MPEG-4 Binary Format for Scenes (BIFS) to exchange their content with other authors, tools, or service providers. XMT is an XML-based abstraction of the object descriptor framework for BIFS animations. Moreover, it respects existing practices for authoring content, such as SMIL, HTML, or Extensible 3D by allowing the interchange of the format between a SMIL player, a Virtual Reality Modeling Language player, and an MPEG player. It does this using the relevant language representations such as XML Schema, MPEG-7 DDL, and VRML grammar. In short, XMT serves as a unifying framework for representing multimedia content where otherwise fragmented technologies are integrated and the interoperability of the textual format between them is bridged.

The MPEG-7 group's objective is to standardize ways of describing different types of multimedia information. The emphasis is on audio–visual content with the goal of extending the limited capabilities of proprietary solutions to identify content by providing a set of *description schemes* and *descriptors* to make various types of multimedia content accessible. In this context, a description scheme specifies the structure and semantics of the relationships between its components, which might be both descriptors and description schemata. A descriptor defines the syntax and the semantics of a distinctive characteristic of the media unit to be described, such as an image's color, a speech segment's pitch, an audio segment's rhythm, a video's camera motion or style, a movie's actors, and so forth. Descriptors and description schemata are represented in the MPEG-7 DDL. The current version of the DDL is based on XML Schema, which provide a means of describing temporal and spatial features of audio–visual media as well as connecting these spatio-temporal descriptions within the media. The DDL also provides the necessary mechanisms for extending and refining existing description schemata and descriptors and to define new schemata or descriptors if required.

## Current problems

Problems exist with using MPEG-7 as the basis for a dynamic media-based knowledge space. First, MPEG-7 is hierarchy centered. This means that a description of data in MPEG-7 is understood as one document that applies a tree structure. The schemata for this document type are fixed and cannot be altered. This linear approach is not astonishing, because efficient access and retrieval was and still is the driving development force of the MPEG-7 standardization effort. However, this approach is far too restrictive; any form of annotation is necessarily imperfect, incomplete, and preliminary, because annotations accompany and document the progress of interpreting and understanding a concept. Graphs, which form the basis of semantic networks, provide better support for carrying out this incomplete task over time.

Related to this problem is the conceptual idea in MPEG-7 of two general description types: complete descriptions (which use the MPEG-7Main as the root element) and partial description units (which use the MPEG-7Unit as the root element). Distinguishing between a complete and fragmental description is purely academic and adds an unnecessary level of complexity.

Another problem is the great number of MPEG-7 schemata—not so much because of their number, which is unavoidable, but because of their interlocked nature, which makes using schemata in isolation difficult.

Finally, it has also become increasingly clear that we need a machine-understandable representation of the semantics associated with MPEG-7 description schemes and descriptors. This representation would enable the interoperability and integration of MPEG-7 with metadata descriptions from other domains. MPEG-7 is currently developing description schemata mainly for the film and broadcasting domain, and to accomplish this, MPEG-7 requires a common understanding of the semantic relationships between metadata terms from different domains. XML Schema, and hence MPEG-7's DDL, provide little support for expressing semantic knowledge, but RDF Schema might. Jane Hunter and Carl Lagoze offer an example for interoperability between application profiles in RDF and XML Schema.[1]

Striving to be a highly interoperable standard among well-known industry standards and other related standards of different domains is a courageous and farsighted step for a group mainly known for its concern with efficient audio–visual coding at the bit level. Moreover, the textual representations in

MPEG-4 and MPEG-7 not only support the current trend in content description toward XML as the accepted standard, but they also point to new ground. Because textual representations allow a symbolic representation of multimedia content by expressing relations between elements—synchronized with the different modalities of multimedia data—it is now possible to model central aspects of how humans try to make sense of complex systems.

So, has the paradigm change in multimedia computing happened yet? Not really, but we're moving in the right direction. The real challenges are still ahead of us—generating and using quality metadata.

It took nearly 30 years of steady infiltration of technological advances in everyday production environments—such as nonlinear video-editing systems, image-editing tools, audio systems, and Web presentation technology—to communicate ideas in forms other than text. And still, the technology follows the strains of traditional written communication by supporting the linear representation of an argument, which results in a final multimedia product of context-restricted content. Thus, we face the paradoxical situation that although there are more possibilities than ever to assist in the creative development and production processes of media, we still lack adaptive environments that can serve as an integrated information space for use in distributed productions, research, restructuring (such as by software agents), or direct access and navigation.

We need systems for authoring media that let people use their creativity in familiar ways and their human actions to extract the significant syntactic, semantic, and semiotic aspects of the media's content to construct descriptions based on a formal language. There is much evidence that manual labor can provide a great deal of useful annotation.[2–4] We also need systems that manage independent media objects and representations for use in many different productions with a potentially wide range of applications.

Yet, if we only had the information gathered during the production of media, including its reuse and modifications, we would still lack knowledge about the material's potential intrinsic meanings. Thus, it is important to make people aware that the notion of a completed work vanishes in such a system and leaves space for a creative and productive cycle, a living environment allowing all sorts of processes. These spaces are for investigation based on an interpreting, associative method rooted in a discourse-oriented collective interpretation of questions that, by following the branches of interdependencies, compare the most diverse theories.

## References


1. J. Hunter and C. Lagoze, "Combining RDF and XML Schemas to Enhance Interoperability Between Metadata Application Profiles," *Proc. 10th Int'l WWW Conf.*, 2001, pp. 456–466.

2. C. Dorai and S. Venkatesh, "Bridging the Semantic Gap in Content Management Systems: Computational Media Aesthetics," *Proc. 1st Conf. Computational Semiotics for Games and New Media* (COSIGN 2001), 2001, pp. 94–99; www.kinonet.com/conferences/cosign2001/program.html (current Jan. 2002).

3. A.T.G. Schreiber et al., "Ontology-Based Photo Annotation," *IEEE Intelligent Systems*, vol. 16, no. 3, May/June 2001, pp. 66–74; http://computer.org/intelligent/ex2001/x3066abs.htm (current Jan. 2002).

4. F. Nack and W. Putz, "Designing Annotation Before It's Needed," *Proc. 9th ACM Multimedia Conf.*, ACM Press, New York, 2001, pp. 251–260; http://acm.org/sigs/sigmm/MM2001/ep/toc.html#Wp1 (current Jan. 2002).


## Language Games for Emergent Semantics


Luc Steels, *University of Brussels AI Lab and Sony Computer Science Lab, Paris*


Every computer scientist knows that we can only process information when the information is somehow represented—there's no computation without representation. Traditionally, human programmers have designed the representations. They select what aspects of the domain are relevant and thus must be made explicit, and they design appropriate data structures that efficiently support the processing required for a task. This works reasonably well, but we need a massive amount of programs these days, making it difficult to keep up. Moreover, users want their programs to adapt to new tasks and a changing world. This raises the question of whether computer systems can develop and adapt representations.

A typical example is Web applications, which must cope with constantly changing information sources (material appears and disappears without any central control) and needs (the Web touches on all aspects of human life and is therefore basically open-ended). Another example is autonomous robots, which must operate in an open-ended and unpredictable world in which new tasks can arise that the designers could not have foreseen.

The origin of representation has been a central topic in AI research from the beginning—it is a problem that human biology has had to solve as well. The question is usually studied under the heading of machine learning and is far from resolved. Indeed, there is a profound paradox.

Computation requires a representation, but how can this computation generate its own representation? A representation casts a frame on the world, but this frame is a strength as well as a limitation. Stepping out of the frame is like jumping out of a hoolahoop while holding it. As Ludwig Wittgenstein put it, "The limits of my language mean the limits of my world."

We can schematically classify efforts to understand the origins of representations into two approaches: induction and selection. I propose a third alternative, which relies on interaction, construction, and communication.

### Induction

The inductive approach is the best known and furthest developed, having been explored in the fields of statistical-pattern recognition,[1] symbolic machine learning,[2] and neural-network research.[3] A large training set must be available, and the inductive process goes over these data to find what is essential and what is contingent. Either the process is supervised, in the sense that it receives feedback about what it needs to learn, or it is unsupervised, in which case it attempts to detect the natural classes or regularities in the data. In the past decade, researchers have developed a wealth of induction algorithms, and many applications have been demonstrated for more compact coding of the data, finding similarities, learning inference rules, data mining, and so forth.

However, some fundamental limitations have come up as well, in the sense that the intervention of a human designer is much greater than hoped for. The designer must assemble an adequate set of training data,

which she must prepare carefully. Often she must choose the outline of the representation to bias the learning algorithm. She must carefully select the learning architecture for the task and domain and set parameters. Pure unsupervised induction often leads to concepts that are irrelevant for the task at hand. For example, a series of real world images might cluster based on the time of the day they were taken rather than on the objects contained in the image.

Many algorithms do not support incremental learning and could even deteriorate in performance when the learning goes on for too long. All of this does not diminish induction's usefulness but suggests that humans might have more up their sleeves.

## Selection

Neo-Darwinian models of genetic evolution and observations of early brain development have inspired an alternative approach to machine learning. This selectionist approach assumes there is a process that generates at random a wide variety of possible representations (alongside the algorithms that use these representations) and a selection process that picks out algorithms and representations best suited for the task.

Again in the past decade, researchers have come up with a wide variety of algorithms and impressive demonstrations of practical applications.[4] It seems that such a selectionist approach is particularly good in optimization, but there is a catch. A human designer once again must heavily interfere to get reasonable results. He must translate the desired target into selection criteria. He must also choose the appropriate operators, which implicitly biases the representations that might develop. Selection is efficient when many solutions can be explored in parallel but is obviously less appropriate when a single agent (such as a robot) or a computer system must individually try out all possibilities. Because selection is based on a parallel search process, it has all the negative characteristics of search, such as the risk of ending up in local minima or long search time when there are no adequate heuristics. Thus, selection is also not the final answer.

## Construction

Suppose you and your family just moved into a new house. Instead of throwing everyone's shoes into one big box, you ordered them, perhaps putting them into separate boxes or drawers with labels, so family members could remember which shoes are in which box. This implies that you not only categorized the shoes but also created an external representation of the categorization to communicate within the group.

This is a typical example of human semiotic activity. Induction couldn't have solved this problem, because there were no concepts to be learned, training sets, or existing labels. Similarly, randomly generating representations would have been an odd way to organize the shoes; imagine the entire family putting shoes into piles, hoping the result would be an effective organization. In reality, three activities must occur: interaction, construction, and communication.

Interaction involves tasks and activities that generate the need for new meanings—in this case, we need to organize the shoes so the family members can easily find their shoes. Construction involves agents that can impose new categories. These categories are chosen in function of the task. For example, the agents could order the shoes based on who wears them, their size, the season in which they're worn, their age, and so forth. These categorizations are not based on natural categories (for example, a grouping based on color would not make sense) but rather on features that will guide the shoes' retrieval. With communication, external tokens associated with the categories intervene—for example, shoes are put into labeled drawers. These communication conventions must be negotiated among those involved in the task—there are no absolute pregiven conventions. Communication is crucial, because it is the motor for testing the concepts' adequacy and for pushing the development of new concepts when there are misunderstandings or task failures.

This intuitive model suggests new ways to approach the problem of emergent semantics. I have performed numerous experiments with teams at the University of Brussels and the Sony Computer Science Laboratory in Paris to better understand this social construction of meaning and to turn it into an operational model.[5] We've learned that this will require several ingredients.

First, we need multiagent systems. If representations result from a collective effort based on interaction and communication between agents, then we cannot restrict ourselves to a single agent or a computer program that is passively receiving a stream of training cases. In addition, the agents need a rich basis of interactions between themselves and the world. In our own work, we are particularly interested in grounding meaning in the real world through a sensorimotor apparatus. So, we need a sufficiently complex body and a rich enough environment, involving other robots as well as humans. In our experiments, we use autonomous robots, such as the Sony AIBO pet robot, as a platform for the agent. These robots are capable of operating in an open environment without narrow prior task definitions.

The second component is construction. Concepts need not arise gradually by stripping away contingent properties of individual cases. Instead, it is possible to take a sensory dimension (say, the objects' size) and introduce a distinction by cutting the space into two distinct regions. One region would correspond to the concept of *small* and the other to that of *large*. If necessary, the dimension could be cut up further into *very small* or *very large* regions, and so forth. More sophisticated construction operators work with multidimensional spaces or with prototypes used through nearest neighbor algorithms.

The third component is communication. Agents should be designed to communicate as part of the task, which means that they must develop symbols to externalize their conceptualizations of reality. Human designers need not define the symbols in advance. Several researchers have now shown abundantly in various experiments that agents can invent new physical tokens (words, grammatical constructions, gestures), using random combinations from an alphabet, and associate these with concepts they want to express or which they can hypothesize in others.[6] It is crucial that the memories of the associations keep a score between a symbol and its meaning, use the symbol–meaning pair with the highest score when they must communicate, and update this score based on feedback in success in the communication. This way, distinctions propagate in the population and become building blocks for more complex representations. They give rise to a culture in which concepts evolve in a memetic process and in which there is a true coevolution of language (external representations) and meaning. Individual agents are engaged in constructing new representations (both internally and externally).

The construction process is collective (due to coordination through language) and incremental, just like evolution by natural selection is collective and incremental. The amazingly rich and ever expanding human conceptual frameworks we see today are the outcome of centuries of incremental construction processes preserved through cultural transmission. Potential applications include the formation and adaptation of ontologies for Web-based agents[6] or evolving dialogs with humanoid robots.[7]

Although much needs to be done to turn these ideas into commonly used technology, we can see the beginnings of a new approach to emergent semantics, complementary to induction and selection.

## References


1. R. Duda, P. Hart, and D. Storck, *Pattern Classification and Scene Analysis*, John Wiley and Sons, New York, 2000.
2. T. Mitchell, *Machine Learning*, McGraw-Hill, New York, 1997.
3. C. Bischop, *Neural Networks for Pattern Recognition*, Oxford Univ. Press, Oxford, UK, 1995.
4. D. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading, Mass., 1989.
5. L. Steels, "Language Games for Autonomous Robots," *IEEE Intelligent Systems*, vol. 16, no. 5, Sept./Oct. 2001, pp. 16–22.
6. L. Steels, "The Origins of Ontologies and Communication Conventions in Multi-Agent Systems," *Autonomous Agents and Multi-Agent Systems*, vol. 1, no. 1, Oct. 1998, pp. 169–194.
7. L. Steels, "Social Learning and Verbal Communication with Humanoid Robots," *Proc. IEEE-RAS Int'l Conf. Humanoid Robots*, IEEE Press, Piscataway, N.J., 2001.


# Emergent Semantics for Ontologies


Alexander Maedche, *University of Karlsruhe*


Tim Berners-Lee coined the vision of a Semantic Web, in which background knowledge on the meaning of Web resources is stored through the use of machine-processable metadata. The Semantic Web should bring meaning to the content of Web pages, and ontologies and metadata have been recognized as the key ingredient for putting the Semantic Web into practice.

In 1999, the World Wide Web Consortium published the Resource Description Framework (RDF) as a recommendation for a framework and data model for metadata representation on the Web. Recently, standardization bodies, in cooperation with different research projects, have been heavily working on formal knowledge representation languages for the Web such as OIL or DAML+OIL, building on RDF and its associated typing system RDF-Schema. These formal languages let us represent complex structured ontologies that impose a formal, machine-processable semantics to the primitives in use.

However, the Semantic Web should also bring added value to human users, not just serving machine needs for understanding content. As human users typically apply context-dependent rules within the process of deriving the meaning for a specific word, it is obvious that the communication process between humans and machines cannot be ensured by formal, machine-processable and understandable semantics alone.

Emergent semantics is a research topic that deals with this aspect. Thus, we must explicitly distinguish between machine-readable and machine-interpretable conceptual structures with a clearly defined, formal machine semantics and the semantics humans assign to conceptual structures in a specific context. Within emergent semantics, we'll investigate and establish means for supporting the interaction between humans and machines for the cooperation and generation of meaning.

## Semiotics and computational semiotics

Many research communities have researched the general problem of how the context-dependent meaning is assigned to a specific word. Here, I refer to semiotics —that is, the study of signs, symbols, and signification, and of how meaning is created. Researchers recently coined the term *computational semiotics*. It involves using and evaluating semiotic theories to analyze, design, and develop computing systems.

In general, there are three levels for understanding a message:

- Syntactic: Which natural language primitives are used? Which meanings may be assigned to these natural language primitives within the system?
- Semantic: What is the meaning of the primitives used within the system?
- Pragmatic: How do humans interpret the natural language primitives? How do humans use natural language primitives for particular purposes?

Humans require words (or general symbols) to talk and to communicate about things. The mapping from words to things is indirect and takes a detour over concepts. The meaning of a given word in context and its reference to a concrete thing is given by a *concept*. The mapping from words to concepts is the result of the human communication process. The meaning triangle in the tradition of Charles Peirce, Ferdinand de Saussure, and Gottlob Frege described by Charles Ogden and Ivory Richards in 1923 explains this relationship.[1] It illustrates that although symbols cannot completely capture the essence of a reference (concept) or of a referent (thing), there is a correspondence between them. The relationship between a symbol and a thing is indirect. The link can only be completed when an interpreter processes the symbol, which invokes a corresponding concept and then links that concept to a referent (thing) in the world. How concepts themselves occur is a matter of discussion among philosophers.[2] Let's focus on intelligent systems and the Semantic Web.

Ontologies try to formalize natural language to enable machine-processable and understandable data. However, as I mentioned, an important aspect is that humans must agree with a specific ontology and its intended conceptualization. Ontology development is a cooperative standardization process, and one crucial point within it is the communication between the different members that should later agree on the defined standard. To support the evolution of a common meaning in the engineering process, we must consider not only formal semantics but also human cognitive structures and the way humans communicate them.

Researchers at the University of Karlsruhe have used two comparatively simple means to do this. First, we introduced an explicit lexical layer for the Web ontology and metadata representation language, RDF(S) (RDF(S) unites RDF and RDF Schema). Thus, we enabled a connection between ontologies and the conceptual system such as that communicated through natural language. We developed the lexical layer by extending RDF Schema, using a new namespace (see http://kaon.semanticweb.org/2001/11/ kaon-language). This pro-

**Simone Santini** is a researcher at the national center for microscopy and imaging research at the University of California, San Diego. His research interests include interactive image and video databases, data models and query models for structured and numerical data (typically image and video features) and temporal data, and multimodal information management. In particular, he is interested in the characterization of the process by which different media, user characteristics, and the circumstances of a query contribute to the emergence of meaning. He received the Laurea degree from the University of Florence, Italy and an MSc and PhD in computer science from the University of California, San Diego. Contact him at ssantini@ncmir.ucsd.edu.

**Frank Nack** is a senior researcher at CWI, currently working in the multimedia and human–computer interaction group. His research interests include video representation, digital video production, multimedia systems that enhance human communication and creativity, interactive storytelling, and media-networked oriented agent technology. He received his PhD from Lancaster University, UK. He is an active member of the MPEG-7 standardization group and is on the editorial board of *IEEE Multimedia*, where he edits the Media Impact column. Contact him at Frank.Nack@cwi.nl.

**Luc Steels** is a professor of artificial intelligence at the University of Brussels and the director of the Sony Computer Science Laboratory in Paris. His research interests in AI include robotics, vision, learning, and natural language, with a focus on the development of computational and robotic models for studying the origins of language. Contact him at AI-lab, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium; steels@arti.vub.ac.be.

**Alexander Maedche** is head of the Knowledge Management Research department at the FZI Research Center for Information Technologies at the University of Karlsruhe. His research interests include knowledge discovery in data and text, ontology engineering, learning and application of ontologies, and the Semantic Web. He received a diploma in industrial engineering and a PhD in applied informatics, both from the University of Karlsruhe. He is a member of the IEEE and GI. Contact him at FZI, Univ. of Karlsruhe, 76131 Karlsruhe, maedche@fzi.de; www.fzi.de/wim.

vides a simple middle ground between a cognitively motivated approach—such as WordNet—and formal semantics.

Second, based on this intermediate layer, we use the implicit semantics contained in communication protocols—that is, in the form of data such as documents and email within the ontology engineering process. In the classical ontology engineering process, the conventional wisdom implicitly contained in this kind of input data is often neglected. Thus, we have developed a means to semiautomatically extract ontologies from different kinds of input data. We provide a short description of the developed ontology learning approach, which builds on the seminal work of text knowledge acquisition done by Doug Skuce and his colleagues in the 1990s.[3]

## Lexical layer for ontologies

In RDF(S), a unique uniform resource identifier defines the ontology's elements. URIs are the Semantic Web architecture's primary element, and they identify resources on the Web such as documents, images, downloadable files, services, and electronic mailboxes. However, URIs are typically neither human readable nor understandable. RDF(S) proposes using so-called *labels* that provide a human-readable version of a resource name. There is no clear definition of how to use these labels—even the labels' language definition falls back to XML syntax without being processable in the RDF data model.

In our work, we pursue a more complex model that lets us assign different kinds of lexical entries to URIs. The distinction between lexical entry and concept is similar to the distinction of word form and so-called synset (synonym sets) used in WordNet. (WordNet was conceived as a mixed linguistic and psychological model about how people associate words with their meaning.) Beside the standard primitive label that is also available in RDF Schema, our extension defines specific lexical entries of URIs such as synonyms, word stems, and so forth.

## Ontology learning

Ontology learning is a bottom-up approach, which starts from a given set of data that reflects the human communication and interaction process. Thus, it considers the discovery of semantics implicitly contained in the existing data that humans have generated by exchanging signs. It includes several complementary disciplines such as machine learning and statistics to support semiautomatic, cooperative ontology engineering.

In our previous work, we introduced a complex framework for ontology learning from different kinds of data.[4] We mainly distinguished between extracting an ontology from scratch and supporting the ontology evolution process by analyzing legacy and application data appearing in different forms. Emergent semantics requires support from both ontology learning steps. First, we can analyze existing results of the human communication process to get a first version of an ontology. Second, an existing ontology used within an application might evolve over time—meanings might change over time to reflect user behavior.

Obviously, the communication process between humans and machines cannot be ensured by formal, machine-processable and understandable semantics alone. Users typically apply context-dependent rules within the process of meaning generation. To establish a consensus within this process or even to formalize the rules humans apply within meaning generation is a future challenge. Upcoming research must deal with the communication documented in texts as well as with communication processes such as ontology engineering sessions performed over the Web. ■

## References

1. 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*, 10th ed., Routledge & Kegan Paul Ltd., London, 1923.
2. P. Gardenförs, *Conceptual Spaces—The Geometry of Thought*, MIT Press, Cambridge, Mass., 2001.
3. D. Skuce et al., "A Logic-Based Knowledge Source System for Natural Language Documents," *Data and Knowledge Engineering*, vol. 1, no. 3, 1985, pp. 201–231.
4. A. Maedche and S. Staab, "Ontology Learning for the Semantic Web," *IEEE Intelligent Systems*, vol. 16, no. 2, Mar./Apr. 2001, pp. 72–79.