Emergent Semantics and Cooperation in Open Systems

[Keynote Abstract]

Tiziana Catarci
Dipartimento di Informatica e Sistemistica "A.Ruberti"
SAPIENZA - Università di Roma
Via Ariosto 25 - 00185 Roma, Italy
catarci@dis.uniroma1.it

Information systems of every organization (ranging from large companies to individual entities) have to handle a variety of information sources, from proprietary ones to information publicly available in web services worldwide. Grasping relevant information wherever it may be and exchanging information with all potential partners has become an essential challenge. Basically, information sharing, rather than information processing, is IT's primary goal in the 21st century. The key point is that now information has to be sharable in an open environment, where interacting peers do not necessarily have a common understanding of the world at hand, as used to be the case in traditional enterprise information systems. Lack of common background generates the need for explicit guidance in understanding the exact meaning of the data, i.e., their semantics. Data semantics is more and more context- and time-dependent, and cannot be fixed once and for all at design time. Identifying emerging relationships among previously unrelated information items (e.g., during data interchange) may dramatically increase their value. Such relationships are the basic ingredients for semantic interoperability that is viewed as an emergent phenomenon constructed incrementally, and its state at any given point in time depends on the frequency, the quality and the efficiency with which negotiations can be conducted to reach agreements on common interpretations within the context of a given task. This type of semantic interoperability is referred to as "emergent semantics" [3, 5].

Software agents have various mechanisms at their disposal for establishing relationships between internal symbols and external meaning. In many cases, humans are responsible for providing the initial semantics. In the simplest case, the natural language vocabulary is used for the local symbols and their relationship with the definition of the notion concerned is left implicit. Often, the hidden assumption is that the local symbol meaning is identified through human cognition. In order to address some of the problems arising when leaving interpretation of the symbol implicit semantics to human cognition, some researchers have proposed to use an explicit, shared reference system for relating sets of symbols. *Ontologies* serve this purpose: they consist of explicit, partial definitions of the intended meaning of symbols for a domain of discourse. Unfortunately, building shared ontologies is a complex process and top-down ontology design, even when done collaboratively, is known not to scale well. Moreover, ontologies are not enough to achieve semantic interoperability. For instance, ontologies are forms of "a-priori" agreements on concepts, and therefore, their use is insufficient in ad-hoc and dynamic situations where the interacting parties did not anticipate all the interpretations and where "on-the-fly" integration must be performed [4]. Indeed, emergent semantics is a global state that should result from the dynamics of local interactions, without any predefined agreement. Such a state cannot be predicted from individual behaviors, nevertheless single interacting peers should be able to analyze feedback from the overall network and infer from such a feedback the reliability of shared context. Given their characteristics, emergent semantics systems are typically peer-to-peer and implemented on top of so-called semantic overlay networks [1]. Research on such systems is going on, still many open problems exist, e.g., global semantic integrity and global consensus; efficiency and scalability; trust, quality, and reputation; automatic construction of local consensus; resource location and identification; uncertain, imprecise, inconsistent, and incomplete information.

An example of systems dealing with semantic interoperability in dynamic open environments, i.e. emergent semantics, is Esteem (Emergent Semantics and cooperaTion in multiknowledgE EnvironMents) [6].

The Esteem approach proposes a comprehensive framework and platform for data and service discovery in P2P systems, with advanced solutions for trust and quality-based data management, P2P infrastructure definition, query processing and dynamic service discovery in a context-aware scenario. The system allows one to access data and services in a simple and effective way, by querying information sources that are similar to the user's interests. Common interests identify semantic communities, which represent semantic affinity between peers emerging in a dynamic and heterogeneous environment. Data and service discovery is performed inside the borders of such communities.

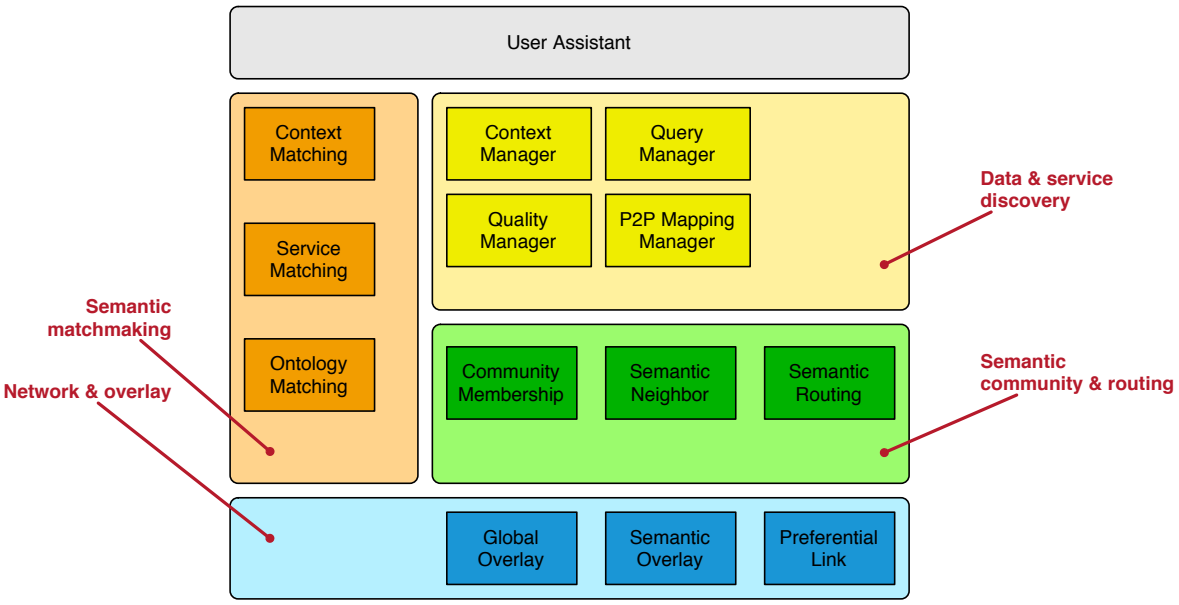

Figure 1: The reference architecture of an Esteem peer

Semantic communities are created and updated in an automatic way by collecting information sources whose contents present high similarity. A threshold-based mechanism allows to establish internal cohesion of community contents, enabling peer aggregation apart of their terminological differences. Semantic communities do not constrain participants to adhere to a global ontology, but compare reference ontologies used by information sources belonging to the same community. In the Esteem system context-aware data and service selection excludes from search results resources that are not accessible from the particular user context. The Esteem system is also in charge of protecting the users from retrieving data and services from untrustworthy information sources.

Esteem relies on an overlay P2P network where i) semantic communities are defined to aggregate peers with similar interests and ii) a probe/search mechanism is adopted to enforce data and service discovery/sharing. An Esteem semantic community sc is defined as a pair of the form $sc = \langle CID, M \rangle$ where CID is the unique Community Identifier that characterizes the community sc and M is the Manifesto, that is the community ontology that describes the common interpretation (i.e., perspective) of the community interests. In Esteem, a semantic community is autonomously emerging, in that it originates from a proposal of a community founder (i.e., a peer) which initiates the community formation through dissemination of an advertisement message that contains CID and M of the emerging community. Each receiving peer p_i autonomously decides whether to join the community on the basis of its level of interest in the received manifesto M. Such a level of interest is computed by invoking an ontology-based semantic matchmaker and by evaluating the semantic affinity between M and the peer ontology of p_i. Furthermore, communities are exploited as a semantic overlay on top of the basic P2P overlay (i.e., the global overlay) in order to enforce effective data and service sharing. In this respect, the probe/search mechanism is used to characterize:

- the discovery phase, based on ontology matching, where probe queries are defined to identify the peers that are capable of providing relevant knowledge with respect to a given topic of interest;
- the sharing phase, based on P2P mapping definition, where standard search queries are issued to point-to-point interact with a previously discovered peer for actual data acquisition and/or service invocation.

In the discovery phase, the joined semantic communities are exploited by a requesting peer for selecting the probe query recipients with the aim of choosing those communities and peers that are most likely to provide relevant results according to the query target. In this context, the semantic matchmaker is invoked to evaluate the relevance of a community with respect to a probe query by comparing the community manifesto against the query content. By collecting probe query replies, a peer evaluates the results and decides whether to perform the sharing phase by directly interacting with the most interesting peers that provided a reply through appropriate search queries with the aim of accessing their data and services.

An Esteem peer is equipped with (all or a subset of): (i) a semantic description of shared data and services, to properly identify its interests, expressed through ontologies; (ii) the representation of context(s) from which the peer accesses data and invokes services; (iii) the representation of quality and trust metadata attached to its data and services (quality profile). When joining semantic communities that share its

interests, the peer also maintains information about joined communities. The Esteem architecture addresses the main requirements of a peer involved in P2P semantic cooperation. As shown in Figure 1, the main components are:

- *Network & overlay.* It is responsible for managing the peer connectivity and for handling incoming and outgoing messages. From the network point of view, the Esteem P2P infrastructure is organized in semantic overlays featuring the semantic communities. In this respect, the network & overlay component is responsible for maintaining the overlays and the associated peer communications.
- *Semantic community & routing.* It is responsible for managing the peer participation in semantic communities and for discovering the semantic neighborhoods of a peer. Furthermore, this component is responsible for providing a semantic routing mechanism to effectively enforce query propagation.
- *Semantic matchmaking.* It is responsible for providing semantic affinity evaluation when comparing different peers' ontological descriptions. This component is invoked by a peer during the discovery phase to identify peers that are capable of providing matching resources (i.e., data, service, context) w.r.t. a given target request. Different techniques are provided by the semantic matchmaking component according to the type of matching resource that is specified in the request. In particular, ontology, service, and context matching techniques are provided by the semantic matchmaker.
- *Data & service discovery.* It is responsible for interacting with the user and for satisfying its discovery requests. In particular, this component provides the functionalities for context and quality/trust management. Furthermore, discovery and sharing functionalities are also addressed in this component through query/answer and P2P mapping management.

Moreover, the system supports the (human) user with an intuitive Web interface that assists him/her in joining the semantic communities that share his/her interests and in identifying data and services he/she is looking for. The Esteem system has been validated with a set of specialists from the health-care domain both to collect system requirements in the first stage of the project and to test the system through a think aloud evaluation technique [2], getting satisfactory results.

1. REFERENCES

[1] A. Crespo and H. Garcia-Molina. Semantic Overlay Networks for P2P Systems. In *Proc. of the 3rd Int. Workshop on Agents and Peer-to-Peer Computing (AP2PC 2004)*, New York, NY, USA, 2004.

[2] A. Dix, J. Finlay, G. Abowd, and R. Beale. *Human-Computer Interaction (3rd Edition).* Prentice-Hall, 2003.

[3] K. Aberer et al. Emergent Semantics Principles and Issues. In *Proc. of the 9th Int. Conference on Database Systems for Advances Applications - DASFAA 2004*, Jeju Island, Korea, 2004.

[4] A. Ouksel. In-Context Peer-to-Peer Information Filtering on the Web. *SIGMOD Record*, 32(3), 2003.

[5] P. Cudrè-Mauroux et al. Viewpoints on Emergent Semantics. *Journal on Data Semantics (JoDS)*, VI, 2006.

[6] The ESTEEM Project Web Site. Emergent Semantics and cooperaTion in multi-knowledgE EnvironMent. http://www.dis.uniroma1.it/∼esteem/.