Emergent Semantics in Distributed Knowledge Management

Carola Aiello[2], Tiziana Catarci[2], Paolo Ceravolo[1], Ernesto Damiani[1], Monica Scannapieco[2], and Marco Viviani[1]

[1] Dipartimento di Tecnologie dell'Informazione- Università di Milano
via Bramante, 65 - 26013, Crema (CR), Italy
{damiani, ceravolo, viviani}@dti.UNIMI.it
[2] Dipartimento di Informatica e Sistemistica - Sapienza Università di Roma,
Via Ariosto 25, - 00185, Roma, Italy
{carola.aiello, tiziana.catarci, monica.scannapieco}@dis.uniroma1.it

Abstract. Organizations and enterprises have developed complex data and information exchange systems that are now vital for their daily operations. Currently available systems, however, face a major challenge. On todays global information infrastructure, 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 business value. This chapter introduce and discuss the notion of Emergent Semantics (ES), where both the representation of semantics and the discovery of the proper interpretation of symbols are seen as the result of a selforganizing process performed by distributed agents, exchanging symbols and adaptively developing the proper interpretation via multi-party cooperation and conflict resolution. Emergent data semantics is dynamically dependent on the collective behaviour of large communities of agents, which may have different and even conflicting interests and agendas. This is a research paradigm interpreting semantics from a pragmatic prospective. The chapter introduce this notion providing a discussion on the principles, research area and current state of the art.

1 Introduction

On todays global information infrastructure, access to information involves interaction with distributed sources. Moreover network agents want to access information efficiently, protecting their sensible information and preserving their autonomy. One of the most important ways of improving the effectiveness of information retrieval and service sharing is by explicitly describing information services semantics. Ontology serves this purpose: it consists of explicit, partial definitions of the intended meaning of symbols for a domain of discourse. Unfortunately, building shared ontology is a complex process and top-down ontology design, even when done collaboratively, is known not to scale well. Also domains are rapidly evolving and the semantics of data cannot be fixed once and for all at design time. This requires to make semantics more and more context- and time-dependent.

Emergent Semantics has been proposed as a solution to the semantic interoperability problem. The general idea is that if we renounce to a centralized control over semantic description, we can improve them by exploiting implicit information emerging during data exchange. For instance repeated downloads are a confirmation of data quality, while the frequency of interactions can define a degree of correlation. This paradigm can be applied in many environments such as human-computer interaction, language games for robot-robot and robot-human communication, scientific databases, e.g. biological data, where data is captured through experiments and subsequently analyzed. In this work, we limited our analysis to distributed knowledge management. This is an environment characterized by the multiple information agents interacting with different level of cooperations. In this context data exchange can be taken as a source of additional information to be used in a process of enrichment of semantic representation. The chapter is organized as follow. In Section 2 we introduce the notion of Emergent Semantics. In Section 3 we discuss the main principles of this research paradigm. In Section 4 we summarize the principal research areas, such as P2P data integration, service discovery or trust and reputation management. Finally, in Section 5 we provide some conclusive remarks.

2 The Notion of Emergent Semantics

The notion of semantics has various definitions in different domains. In the domain of Programming Languages, semantics basically refer to rules which relate inputs and outputs [74]. In logic, semantics is the Kripke's theory of truth, usually expressed by a a set-theoretic interpretation [62]. On the web, semantics is often intended as the metadata used for annotating resources, in this case the term "equivalent semantics" designates for equivalent annotation imposed on different resources, using a common vocabulary, as for instance in [20]. For some applications in the field of image retrieval the semantic interpretation of the image content is available as the result of an automatic or semi-automatic image analysis process, applied to images belonging to specific domains and described in advance [63]. Here the semantics of a resource is a typical pattern related to that resource. As a least common denominator, we can characterize semantics as a relationship or a mapping established between an information items syntactic structure and some external domain.

The Emergent Semantics approach consolidates the local semantics held by autonomous information agents into a global, population-wide semantics that results from the continuous interaction of the agents among themselves. The large-scale structures emerging from these continuous interactions dynamically provide meaning to the local symbols. Semantics constructed incrementally in this way is called Emergent Semantics (ES). This is a research paradigm interpreting semantics from a pragmatic prospective. More complete presentations of this paradigm can be found in [22], [2], [71].

3 Principles and main features of the Emergent Semantics Paradigm

Emergent semantics is the form of semantic interoperability viewed as an emergent phenomenon [2] constructed incrementally among data sources in a bottom-up, semi-automatic manner without relying on pre-existing, global semantic models. In such a dinamic scenario, global information is highly evolutionary: documents of already existing sources may be updated, added or deleted; new sources and services may appear and some may disappear (definitively or not). At any given point in time, the state of the semantic interoperability 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. The main principles and features of the emergent semantics paradigm can be summarized as follows:

- *Semantic agreement.* Two or more peers need to establish a semantic agreement, that is to share the meaning of a model, like a conceptual model, or other relevant meta-data information to the task at hand.
- *Agreement negotiation.* Agreements are negotiated between peers, and are dynamically built and maintaned. Such agreements constitute the basis for communication between peers which is realized in terms of message exchanges. Negotiations of semantic agreements are performed on local basis, that is a peer directly contacts all the peers it wants to communicate with. In this way the number of interactions is greater than the one characterizing environments that involve third parties with the role of mediating the communication. However, though the agreements are built on a local basis, a global agreement is nevertheless obtained as a result of their propagation. This is, in a sense, the real essence of the emergent semantics paradigm: it is not necessary to contact all the network, in fact the communication can be incrementally realized by exploiting semantic agreements. Agreements are established on the basis of the peer's specific goals; hence for each distinct goal a peer could establish distinct agreements.
- *Self organization.* The emergent semantics paradigm relies, at the end, on a self-organization process. Self-organization processes have been studied in many scientific disciplines such as physics or biology for a long time and are recently being investigated in computer science as well. Under this perspective, emergent semantics is another application of this successful principle in the computer science field.

4 Principal Research Areas

Important efforts in many research areas are needed to achieve semantic interoperability by the emergent semantics paradigm. In this section we detail the principal involved research areas : *(i)* P2P Data Integration, *(ii)* Service Discovery and *(iii)* Trust and Reputation.

4.1 P2P Data Integration

Information systems have been characterized by a multitude of autonomous, heterogeneous information repositories. The problem of how to provide transparent access to heterogeneous information sources while maintaining their autonomy is a time-honored one. Information integration systems typically provide a uniform query interface to a collection of distributed and heterogeneous information sources, giving users or other agents the illusion of querying a centralized and homogeneous information system. As such, they are considered as mediation systems between users and multiple data sources which can be syntactically or semantically heterogeneous while being related to the same domain. Existing mediator-based information systems can be distinguished according to: (1) the type of mappings between the mediated schema (on global schema) and the schemas of the sources: there exist basically two approaches for such mappings, the Global As View (GAV) and the Local As View (LAV). The global-as-view approach describes the global schema as a view over all local schemas, whereas the local-as-view approach describes each local schema as a view over the global schema; (2) the languages used for modelling the global schema and the source descriptions and (3) the expressivity of the global schema. Starting from this main thread, a recent research line aims to integrate data in a peer-to-peer (P2P) environment and is identified as peer data management or peer-to-peer data integration. A P2P system is characterized by a structure constituted by various autonomous nodes that hold information and that are linked to other nodes by means of mappings. No global ontology a peer can refer to is actually available. The goal of P2P data integration is to provide unified access to this set of heterogeneous data sources. The lack of any agreed-upon global schema or ontology makes it very difficult for the participating parties in the system to reach a global consensus on semantic data.
In the following we describe the principal research issues related to P2P data integration: *(i)* Formal Semantics, *(ii)* Ontology Matching, *(iii)* Query Processing and *(iv) Data Quality* .

Formal Semantics Initial approaches rely on some pre-defined corpus of terms serving as an initial context for defining new concepts [41] or make use of gossiping and local translation mappings to incrementally foster interoperability in the large [1]. However, there is still a fundamental lack of understanding behind the basic issues of data integration in P2P systems. Indeed, since no single actor is in charge of the whole system, it is unrealistic to assume restrictions on the overall topology of the P2P mappings [42, 31]. Hence, one has to take into account the fact that the mappings may have an arbitrary structure, possibly involving cycles among various nodes. This needs to be addressed both from the point of view of modeling the system and characterizing its semantics (see, e.g., [17, 42] for a first order logic characterization and [13, 14] for an alternative semantics proposal) and from the point of view of computing answers to queries posed to a peer. Query answering has difficulties that arise from the necessity of distributing the overall computation to the single nodes, exploiting their local processing capabilities and the underlying technological framework. A recent proposal highlighting the peculiarities of P2P data integration systems is [37].

Ontology matching Dynamic ontology matching techniques can be used as the basis for driving the negotiation of agreements in order to discover the mappings between concepts of different peer ontologies and maintain them.

The general goal of ontology matching techniques is to compare different ontological descriptions for finding concepts that have a semantic affinity. A survey of ontology matching techniques is provided in [70], where formal and heuristic approaches are classified. The former are founded on model theoretic semantics and automated matching techniques [52, 10, 38], while the latter are based on the idea of guessing mappings that may hold between similar graph structures through a combination of analysis, matching, and learning techniques [27]. In [16], heuristic techniques for performing dynamic ontology matching in open, peer-based networked contexts are proposed. Peculiar features of these techniques are the capability of performing the matching process in a flexible and adaptable way, by dynamically configuring the matching algorithm with the most appropriate matching model for the specific matching case. This is done by taking into account the level of richness of ontological knowledge description as well as the requested degree of mix between the importance of linguistic and contextual features of concepts in the evaluation of their matching degree.

Query Processing A P2P data integration system consists of a set of (physical) peers. Each peer has an associated schema that represents its domain of interest. Some peers store actual data, and describe which data they store relative to their schema; the stored data does not necessarily cover all aspects of the peers' schema. Peers are linked through peer mappings. A mapping is the the medium for exchanging data and reformulating queries among different schemas; in particular, the mapping defines the overlapping parts of acquaintances' schemas. Peer mappings describe the semantic relationship between the schemas of pairs (or small sets of) peers. Given a query over a peer P_i the system will use the peer mappings to reformulate the query over the schemas of the neighboring peers. Typically, when a peer joins a P2P data integration system, it will supply mappings, as is most convenient, to a small number of peers. Semi-automatic schema mapping techniques have been investigated in [64, 25, 26].

The key step of query processing in a P2P data integration system is reformulating a peer's query over other peers on the available semantic paths. Broadly speaking, the P2P data integration system starts from the querying peer and reformulates the query over its immediate neighbors, then over their immediate neighbors, and so on. Whenever the reformulation reaches a peer that stores data, the appropriate query is posed on that peer, and additional answers may be found. Since peers typically do not contain complete information about a domain, any relevant peer may add new answers.

Among the principal approaches to query processing in P2P systems we cite Piazza [42], Hyperion [50] and PeerDB [61].
In Piazza two types of mappings, *peer* and *definitional* mappings are defined, and used for performing query answering. Peer mappings describe data within the stored relations (generally with respect to one or more peer relations); definitional mappings are instead between the schemas of the peers. This approach is alternative to the one used by the Hyperion project [3]. Hyperion mappings rely on the usage of mapping tables that store the correspondence between values. As outlined in [50, 51], mapping tables are often the result of expert knowledge and are manually created by domain specialists. However, mechanisms to partially support the task of mapping discovery can be used: new mappings can be inferred from already existing mappings present in the mapping table.
PeerDB [61] is another P2P-based system for distributed sharing of relational data. Similar to Piazza, PeerDB does not require a global schema, but it doesn't use schema mappings for mediating between peers. Instead, PeerDB employs an information retrieval based approach for query reformulation. In their approach, a peer relation (and each of its columns) is associated with a set of keywords. Given a query over a peer schema, PeerDB reformulates the query into other peer schemas by matching the keywords associated with the two schemas. Therefore, PeerDB does not have to follow semantic paths to reach a distant peer. The resulting reformulated queries in PeerDB may not be semantically meaningful, and ultimately the system requires user input to decide which queries are to be executed.
Besides mappings, also routing indices can be used for propagating queries in the system [21, 43], that provide aggregated information about records that are retrieved in the query processing phase. Hence, such indexes can be used for query optimization purposes.
A further approach for indexing data in peer-to-peer systems is based on communities. A community is built on the basis of the similarity of the schema (or the schema mappings) hold by peers. Query processing benefits from the fact that if a query is posed on a certain peer than members of its community will also have similar data. An example of such an approach is provided in ESTEEM [30].

Data Quality In peer-to-peer environements, where information is exchanged between heterogeneous data sources, the quality of data exchanged and provided by different peers is extremely important. A lack of attention to data quality can imply data of low quality to spread all over the system.
In [56], the problem of the quality of web-available information has been faced in order to select data with high quality coming from distinct

sources: every source has to evaluate some pre-defined data quality parameters, and to make their values available through the exposition of metadata.
When considering the issue of exchanging data and the associated quality, a model to export both data and quality data needs to be defined. Some conceptual models to associate quality information to data have been proposed that include an extension of the Entity-Relationship model [73], and a data warehouse conceptual model with quality features described through the Description Logic formalism [44]. Both models are thought for a specific purpose: the former to introduce quality elements in relational database design; the latter to introduce quality elements in the data warehouse design. In [69], a model for associating quality metadata to data exported in the context of a cooperative and distributed system is described.
An important step for quality-aware data integration is the assessment of the quality of the data owned by each peer. For this task, some of the results already achieved for traditional systems can be borrowed, such as record linkage techniques or data cleaning tools [5].
Data integration must take into account data sources heterogeneity. As described in [75], when performing data integration two different types of conflicts may arise: semantic conflicts, due to heterogeneous source models, and instance-level conflicts, due to what happens when sources record inconsistent values on the same objects. The Data Quality Broker described in [30] is a system solving instance-level conflicts. Other notable examples of data integration systems within the same category are AURORA [75] and the system described in [68]. AURORA supports conflict tolerant queries, i.e. it provides a dynamic mechanism to resolve conflicts by means of defined conflict resolution functions. The system described in [68] describes how to solve both semantic and instance-level conflicts. The proposed solution is based on a multidatabase query language, called FraQL, which is an extension of SQL with conflict resolution mechanisms. Similarly to both such systems, the Data Quality Broker supports dynamic conflict resolution, but differently from them the Data Quality Broker relies onto quality metadata for solving instance-level conflicts. A system that also takes into account metadata for instance-level conflict resolution is described in [33]. Such a system adopts the ideas of the Context Interchange framework [11]; therefore, context dependent and independent conflicts are distinguished and accordingly to this very specific direction, conversion rules are discovered on pairs of systems.
The ESTEEM architecture [30] is an example of an emergent semantics system with data quality support. In the Esteem architecture the data quality profile module involves the computation of data quality metrics on the peer data that are available to other peers. More specifically, each peer has the possibility of associating quality metadata to the exported data (at value level). Such metadata represent data quality measures corresponding to some specific quality dimensions. Metrics for the most common quality dimensions (column completeness, format consistency, accuracy and internal consistency (see [6] for the definition of such metrics) are already implemented and the model is ready to be extended

to other dimensions. Besides the data quality profile module, a quality manager module is invoked during query processing in order to exploit quality metadata and to take data inconsistencies into account. More specifically, it is assumed that data can exhibit key-level conflicts [7]. This implies that a record linkage step must be performed in order to provide answers to user queries.

4.2 Service Discovery

The emergence of semantics is a key issue to enforce timely discovery and dynamic composition of distributed services. Recently, many organizations have heavily invested in Web Service technologies and, as a consequence, a growing number of services is being made available. Service proliferation over the Web has been facilitated by the development of several standards, like WSDL for service description, UDDI for service registry, SOAP for message exchange and BPEL4WS for service orchestration.

The discovery of services is the most important functionality in distributed and service-oriented environments. Standards like UDDI or WSDL support description of services and discovery functionalities from a syntactic perspective. But the major problem remains: the semantics of the service description. Often, the same services are described by users and service providers in different ways. As services and their descriptions are evolving quickly responding to market changes, it is in general impossible to keep up with all requirements in time. A decentralized discovery service exploiting emergent semantics approaches to extend the standards in a controlled way and distribute the changes among the peers appears as a concrete possible solution. With such an approach, a peer could "learn" about new descriptions and mappings incrementally departing from existing standards used for bootstrapping the process. Modern approaches for service discovery have to address the treatment of dynamical aspects both with respect to the continuous addition and removal of services in a highly variable environment and with respect to different contexts in which a service could be invoked [15]. Advanced techniques and tools for enabling semantic service discovery are therefore highly desired and required. In particular, it is necessary that services are described in a formal way and service semantics is well captured. In the literature, ontology-based approaches are being developed to exploit the benefits of the ontology technology, such as inferencing, in the context of service discovery. In the Semantic Web, the ontology description languages OWL [24] and OWL-S [19] have been recently proposed. Service description is composed by a service profile (what the service does), a service model (how the service works) and a service grounding (how to invoke the service).

The Semantic Web Services Initiative (SWSI, see www.swsi.org) relaxes the constraint of using a description logic formalism for defining service workflow, and uses a first-order logic based language. In the Unified Problem Solving Method Development Language (UPML) framework [34] logical expressions defined in goals, mediators, ontologies and Web Services are expressed using frame logic. UPML distinguishes between domain models, task models, problem solving methods and bridges, and is also the basis of the Internet Reasoning Service (IRS) [57], a knowledge-based approach to Semantic Web Services. Domain models are effectively the domain ontology, while task models provide a generic description of tasks to be solved. Problem solving methods provide implementation-independent descriptions of tasks, while the bridges map between the various components.

Service matchmaking has been addressed by several approaches in literature: given a request R and a set of advertisements S, the matching procedure must return the set of advertised services that match better with R, possibly ranked with respect to their level of matching (if it can be evaluated). In most approaches the starting point is the UDDI registry, where service descriptions are published; UDDI registry offers searching functionalities that use traditional keyword-based techniques, featured by low precision and recall. To provide semantic matching between service descriptions, some approaches consider concept definitions within ontologies (concept-based techniques). In [58] a framework for semi-automatically marking up Web service descriptions with ontologies is proposed with algorithms to match and annotate WSDL files with relevant ontologies; domain ontologies are used to categorize Web services into domains. The use of ontologies enables service matchmaking in the discovery process. In fact, the elements used for service capability description refer to concepts that can be properly defined and semantically related in domain ontologies. Semantic relationships between concepts are then exploited to establish the type of matching between advertisements and requests. Dynamic discovery of distributed services is based on semantic interoperability. In [76] a service ontology specifies domain concepts with a set of synonyms to allow a flexible search and a set of service classes to define the properties of services, its attributes and operations. In [12] a new technique for Web service discovery which features a flexible matchmaking by exploiting DAML-S ontologies is proposed. In [49] a Web Service Modeling Ontology (WSMO) is expressed by using the formal F-Logic language to describe various aspects related to Semantic Web Services. They start from the Web Service Modeling Framework (WSMF)[35], that consists of four elements: ontologies that provide terminology used by other elements (concepts, axioms, relations and instances), goals of Web Services (by means of pre- and post-conditions), Web Service description (non functional properties, choreography and orchestration aspects) and mediators which bypass interoperability problems.

4.3 Trust and Reputation

As outlined in previous sections, today's Web infrastructure is increasingly used for semantics-driven access to services and resources. This problem is twofold. The first aspect is related to information retrieval, and can be adressed by intelligent search and selection techniques. The second is deciding which among many sources is most reliable and it is usually presented by the notions of *trust* or *reputation*.

In human society, trust and reputation are social knowledge allowing to evaluate which agents can be considered as a reliable sources of information or services. In computer science Trust is not a new research topic in itself; however even, if there is a rapidly growing literature on the theory and applications of trust in different kind of settings, there is a considerable confusion around the terminology used to describe them. In fact, depending on the area where the concept of trust is used – security and access control in computer networks, reliability in distributed systems, game theory and agent systems, and policies for decision making under uncertainty – it varies in these different communities in how it is represented, computed, and used.

The concept of trust is often connected to the mechanism that verify the identity of a remote source of information; in this context, it is investigated it is in association with signatures and encryption mechanisms, whose purpose is to provide protection against malicious parties. The policies used to manage authorizations, allowing to distinguish trusted and untrusted parties, are defined a-priori by a central authority. In a community, however, trust in strictly connected to the notion of relationship among parties. In distributed community-oriented scenarios, for evaluating the reliability of resources we need to deal with the notion of relationship between a *trustor*, the subject that trusts a target entity, and a *trustee*, the entity that istrusted. The formalization of this notion can significantly improve the quality of the retrieved resources. As stated in [53] and [47], trust can be an important factor in decision-making, because it forms the basis for allowing a trustee to use or manipulate resources owned by a trustor or may influence a trustor's decision to use a service provided by a trustee. According to the Emergent Semantics approach the subjectivity of knowledge is seen as a potential source of value and local misalignment is used as a way for improving and evolving semantic mappings.

Defining Trust *Trust* is a complex subject relating to belief in the honesty, truthfulness, competence, reliability, etc., of the trusted person or service. There is no consensus in the literature on what trust is and on what constitutes trust management, because the term trust is being used with a variety of meanings [55]. Many researchers have recognized the value of modeling and reasoning about trust computationally; however, here is no entierly shared notion of trust nature as many authors in the field have noted, the meaning of trust as used by each researcher differs across the span of existing work.

In [46] two definitions of trust are introduced: *reliability trust* and *decision trust* respectively. The first one is introduced by means of the Gambetta definition of trust [36] as "the subjective probability by which an individual, A, expects that another individual, B, performs a given action on which its welfare depends". This definition includes the concept of *dependence* on the trusted party, and the *reliability* (probability) of the trusted party as seen by the trusting party.

However having high (reliability) trust in a person in general is not necessarily enough to decide to enter into a situation of dependence on that person [32]. Jøsang et al. introduces the definition inspired by McKnight

& Chervany [55] where "decision trust" is "the extent to which one party is willing to depend on something or somebody in a given situation with a feeling of relative security, even though negative consequences are possible".

An alternative definition given in Mui et al. [59], which refers to past encounters, and may be thought as *reputation-based* trust, described as "a subjective expectation an agent has about another's future behavior based on the history of their encounters".

Another interesting definition affirms that trust is "the firm belief in the competence of an entity to act dependably, securely, and reliably within a specified context" (assuming dependability covers reliability and timeliness) [39].

The close relationship between trust and belief is emphasized by the definition by Olmedilla et al. [60], which refers to actions and not competence like the previous one: "Trust of a party A to a party B for a service X is the measurable belief of A in that B behaves dependably for a specified period within a specified context (in relation to service X)".

Depending on the environment where the notion of trust must be applied the suitable definition can spotlight different features. In a distributed environment one of the main features to be considered is the "dynamic nature" of trust, as discussed in [72]. As time passes, the trust one entity has in another might not stay the same. In particular trust can change depending on the *experience* that a trustor has about a trustee, and this experience is always related to a *context*. Another important aspect is the difference between trust and reputation. While the concept of reputation refers to a perception that is generally said or believed about an agent; trust ultimately is an individual phenomenon that is based on various factors or evidence, some of which carry more weight than others. The difference between trust and reputation can be illustrated by the following statements [46]: (a) "I trust you because of your good reputation" and (b) "I trust you despite your bad reputation".

A distributed system relaying on the notion of trust must support two important tasks: learning reputation and reasoning on trust. An agent can learn reputations interacting with other agents and aggregating trust evaluations of other agents. While the action of reasoning on trust describes the process in which an agent integrates the reputations from other agents, with a trust model (public or private) and its own beliefs, to update its local trust model.

Modeling Trust In the previous Section we have introduced the definition of trust. Now we focus our discussion on how realizing this notion can be supported by a model. Models for computing trust can belong to two categories: (i) *policies-based* models or (ii) *reputation-based* models [8]. The definition of "hard evidence" used in policies opposed to the subjective estimation of trust used in reputation systems, as appear in [4], reflect the difference between the term *hard security* used for traditional mechanisms like authentication and access control, and *soft security* for social control mechanisms in general, of which trust and reputations systems are examples. The difference between these two approaches was first described by Rasmusson & Jansson in 1996 [66].

Policies-based models describe the conditions necessary to obtain trust, and can also prescribe actions and outcomes if certain conditions are met. Policies frequently involve the exchange or verification of credentials, which are information issued (and sometimes endorsed using a digital signature) by one entity, and may describe qualities or features of another entity. In this field the terms authorization and authentication are often connected to trust. *Authorization* can be seen as the outcome of the refinement of a more abstract trust relationship. We define authorization as a policy decision assigning access control rights for a subject to perform specific actions on a specific target with defined constraints. *Authentication* is the verification of an identity of an entity, which may be performed by means of a password, a trusted authentication service, or using certificates. There is then an issue of the degree of trust in the entity that issued the certificate. Note that authorization may not be necessarily specified in terms of an identity. Anonymous authorization can be implemented using capabilities or certificates

Reputation-based systems model an assessment on the history of interactions with or observations of an entity, either directly with the evaluator (personal experience) or as reported by others (recommendations or third party verification). How these histories are combined can vary, and recursive problems of trust can occur when using information from others (i.e., can I trust an entity's recommendation about another entity?).

At a basic level, both credentials and reputation involve the transfer of trust from one entity to another, but each approach has its own unique problems which have motivated much of the existing work in trust.

Trust Research Classification Due to the growing interest about trust and the resulting growing corpus of literature on it, there is no shared taxonomy of trust research. However some survey literature does exist approaches have been divided in different areas in literature. In particular [4] organizes trust research in four major areas:

1. *Policy-based trust.* Trust is established simply by obtaining a sufficient amount of credentials pertaining to a specific party, and applying the policies to grant that party certain access rights. The recursive problem of trusting the credentials is frequently solved by using a trusted third party to serve as an authority for issuing and verifying credentials.
2. *Reputation-based trust.* History of an entity's actions/behavior is used to compute trust, and may use referral-based trust (information from others) in the absence of (or in addition to) first-hand knowledge.
3. *General models of trust.* Trust models are useful for analyzing human and agentized trust decisions and for computable models of trust operational. Work in trust modeling describes values or factors that play a role in computing trust, and leans more on work in psychology

and sociology for a decomposition of what trust comprises. Modeling research ranges from simple access control polices (which specify who can be trusted when accessing data or resources) to analyses of competence, beliefs, risk, importance, utility, etc.
4. *Trust in information resources*. Trust is an an increasingly common theme in Web related research regarding whether Web resources and Web sites are reliable. With the advent of the Semantic Web, new work in trust is harnessing both the potential gained from machine understanding, and addressing the problems of reliance on the content available in the web so that agents in the Semantic Web can ultimately make trust decisions autonomously. Provenance of information is key to support trust decisions, as is automated detection of opinions as distinct from objective information.

In his short survey Griffiths [40] divides trust researches in three areas:

1. *Security-oriented trust*. A mechanism for ensuring security, encompassing issues of authentication, authorization, access control, privacy.
2. *Service-oriented trust*. A mechanism for achieving, maintaining, and reasoning about quality of services and interactions.
3. *Socially-oriented trust*. A social notion for modeling and reasoning about the relationships between agents, influenced by social science, psychology or philosophy.

In addition to the taxonomies given by [4] and [40], in [65] trust research is categorized according to the definitions of *individual-* or *system-level* trust; in the former, individual agents model and reason about others, while in the latter agents are forced to be trustworthy by externally imposed regulatory protocols and mechanisms (this category includes the already described area of policy-based trust). Finally, depending on the fact that agents either trust others directly based on their experiences, or base their trust on the recommendations of others, models can divided in *direct-trust-based* or *recommendation-based*.

Trust and Emergent Semantics As seen in Section 3, in an ES setting, the notion of uncertain knowledge is crucial. In [29] authors analyze Knowledge Management Systems describing the shared knowledge distinguish to two dimensions: (i) the number of domains involved and (ii) the number of conceptualization used in order to describe these domains. While in a centralized system domains and conceptualizations are usually in a $1:1$ relationship, and if multiple domains are taken into account a single conceptualization is used ($n:1$); distributed systems have $1:n$ or $n:m$ relationship. This simple remark clearly shows how the mapping among the representations of different peers are subject to uncertain knowledge. Trust can be one of the ways for modeling some aspects related to uncertain knowledge. Reconsidering the discussion proposed in Section 3, we can see that trust can have an impact on many typical actions of peer-to-peer environments. Trust related P2P actors include:

- *Grouping*: a group of peers that agree to be considered as a single entity, at least with respect to a given task, by other peers. To become a member of a peer must provide some credential that can vary

from proving to manage some specific type of resource, to providing a token.
- *Discovery*: is a mechanism that allows the user to discover resources in the peer to peer network.
- *Query Propagation*: when a provider receives a query, it might propagate that query to another provider it considers expert on what it believes is the meaning of the request. In order to decide where to propagate a query a peer has two possibilities: (i) a proximity criteria, i.e. the query will be sent to known peers (i. e. by using the discovery functionality) and selection will be done according to a quantitative criteria (number of peers, number of possible re-routings hops-, etc.); this way peers or providers that are not directly reachable by the seeker or that have just joined the system, can advertise their presence and contribute to the resolution of the query; (ii) a semantic criteria: if the provider computes some matching between a query and concepts in its own context, the query resolution mechanism might look for addresses of other peers that have been associated to the matching concept. Here propagation is done on the base of explicit trust since the provider defines other peers as "experts" on the query topic.

Projects and Applications In this section we discuss the state of the art of projects and applications relying on ES for managing Trust in Distributed Knowledge Management. Several projects adopt an approach that can be associated to the ES principles. For example in [9] authors state that Knowledge Management systems should be designed in order to support the interplay between two qualitatively different processes: the autonomous management of knowledge of individual groups and the coordination required in order to exchange knowledge among them. The authors introduces the notion of a *proximity* criteria, that is related to semantic closure and trustworthiness among peers. Another project directly focusing on Distributed Knowledge Management is SWAP, [28]. Here the notion of Trust is adopted in order to rate the resources shared in a semantic-based peer-to-peer system. However this work does not propose a process for managing the resource rating. A first work discussing architectures and functions form managing aggregation and distribution of reputations on semantic is [45]. In [67] the same problem is addressed using path algebra. In [18] the authors provide a algorithm for aggregating rating on trustworthiness. No mature system implemented on these algorithms exist right now.
Enlarging our attention to the notion of imprecise knowledge another group of researches is available. Here again we note the lack of implemented solutions. In [48] authors provide a preliminary discussion on the role of belief and trust in managing emergence in the context of information and service semantics. A taxonomy of belief level is proposed and a measure of data quality based on belief and trust is scratched. In [1] the issue of managing semantic interoperability among data sources in a bottom-up, semi-automatic manner without relying on pre-existing, global semantic models is addressed. A solution is proposed based on an

heuristic focusing on the cyclic exchange of local mappings. This work strongly underlines the need for a theoretical progress on the uncertainty reduction problem in a distribute environment. A contribution on this line is given in [54] where authors discus how fault tolerance can be managed in a system where trustworthiness on data is constructed by computing a reputation chain. Also in [23] a solution is proposed based on the use of decentralized probabilistic models to reduce uncertainty on schema mappings.

5 Conclusions and Open Issues

In a distributed environment of information agents such as in the Semantic Web or in peer-to-peer systems, where information is spread over heterogeneous sources and no global ontology nor centralized control are available, the aim is to enable agents to interoperate irrespective of the source of their initial/local semantics. Semantic interoperability is a crucial element for making distributed information systems usable.

In this chapter we have presented the emergent semantics paradigm, its features and the principal involved research areas. Emergent semantics consolidates the local semantics of autonomous information agents into a global, population-wide semantics that results from the continuous interaction of the agents among themselves, for this, emergent semantics has been proposed as a solution for the semantic interoperability problem.

Of course, as discussed in previous sections, a lot af open issues need to be addressed by future research. Major challenges to be addressed in future researches are:

- A robust theoretical foundation to advanced methodologies for evolvable semantics representation.
- Developing proof-of-concept knowledge management systems for information-bound organisations and communities, capable of extracting actionable meaning from social interaction patterns.

At first glance, the second challenge might seem narrower than the first one, but it has become clear that emergent semantics applications are legion, so that extraction and representation must be somewhat tunable to domain-specific requirements. In particular focusing on the theoretical foundation researches have to explores the notion of *Incomplete Knowledge*. In a large-scale distributed environment of autonomous agents, information and information needs can no longer be expressed concisely, as expected by database and semantic web technologies, but have to deal with numerous sources of knowledge that can be inconsistent or can express uncertain information. Also inconsistency can insist on global level but not on local level, and this two dimensions must be managed by a composition not effacing disalignment of local sources.

In distributed environments, several qualitatively different sources of uncertainty have to be dealt with as well. Besides uncertainty about users information needs, a comprehensive approach must deal with diverse uncertainty types such as uncertainty on knowledge conceptualizations, uncertainty on metadata assertions and uncertainty

on trustworthiness of the information providers. Current reasoning techniques for handling uncertainty have been developed for isolated problems. It is a well-known fact that complete, probabilistic reasoning is as computationally intractable as reasoning in full first order logic is. A number of other formalisms are available for reasoning under uncertainty, such as Bayesian networks or possibilistic and fuzzy logic. Trust is another important notion related with uncertainty and formalized according to different approaches. Also there are various application domains where no definite semantics can be attached to facts captured in the real world. This is the case for scientific databases, e.g. biological data, where data is captured through experiments and subsequently analyzed to guess what is the phenomenon that may actually be materialized into that data. Another example is forensic science, where investigators may try to understand what is the situation behind the factual evidences that have been collected here and there. Potential semantics are associated by formulating hypotheses, which are then tested to see if they can be supported by the collection of available data. If yes the running hypothesis can be confirmed. If not another hypothesis will be formulated and tested, till a plausible semantics is found. Semantics emerges gradually, thanks to a reformulation and test cycle rather than agreement between different agents. In general it is important to stress that ES approach needs a common abstraction framework for reasoning under uncertainty, handling complex conditional relationships between various sources of uncertainty and their models.
Another way to face the problem is to focus on the notion of data quality. Data stored electronically is usually affected by a number of quality problems, ranging from poor accuracy at a syntactic level (data is wrong, for example as a consequence of errors during data collection or manipulation), to various forms or inconsistency, both in a single source or across multiple sources, to problems related to their currency, and so on. This is true also in classical data integration settings, but the risk of poor data quality becomes even more critical in emergent semantics settings. Due to the open nature of these networks, no restriction at all can be enforced on the quality of the data shared by the peers. Thus, it is essential that data quality control be enforced both through the trust and privacy mechanisms, and with proper measures at query processing time that explicitly take quality into account.

6 Acknowledgments

This work was partly funded by the Italian Ministry of Research under FIRB contract `RBNE05FKZ2_004` (project TEKNE).

References

1. Karl Aberer, Philippe Cudré-Mauroux, and Manfred Hauswirth. The Chatty Web: Emergent Semantics Through Gossiping. In *Proceed-*

ings of the 12th International World Wide Web Conferenc e, 2003.
2. Karl Aberer, Philippe C. Mauroux, Aris M. Ouksel, Tiziana Catarci, Mohand S. Hacid, Arantza Illarramendi, Vipul Kashyap, Massimo Mecella, Eduardo Mena, Erich J. Neuhold, and Et. Emergent semantics principles and issues. *Lecture Notes in Computer Science*, pages 25–38, March 2004.
3. Marcelo Arenas, Vasiliki Kantere, Anastasios Kementsietsidis, Iluju Kiringa, Renée J. Miller, and John Mylopoulos. The hyperion project: from data integration to data coordination. *SIGMOD Record*, 32(3):53–58, 2003.
4. Donovan Artz and Yolanda Gil. A survey of trust in computer science and the semantic web. To appear in Journal of Web Semantics, 2007.
5. C. Batini and M. Scannapieco, editors. *Data Quality: Concepts, Methodologies, and Techniques*. Springer, 2006.
6. C. Batini and M. Scannapieco, editors. *Data Quality: Concepts, Methodologies, and Techniques (chapter 2)*. Springer, 2006.
7. C. Batini and M. Scannapieco, editors. *Data Quality: Concepts, Methodologies, and Techniques (chapter 5)*. Springer, 2006.
8. Piero Bonatti and Daniel Olmedilla. Driving and monitoring provisional trust negotiation with metapolicies. In *POLICY '05: Proceedings of the Sixth IEEE International Workshop on Policies for Distributed Systems and Networks (POLICY'05)*, pages 14–23, Washington, DC, USA, 2005. IEEE Computer Society.
9. M. Bonifacio, R. Cuel, G. Mameli, and M. Nori. A peer-to-peer architecture for distributed knowledge management. In *In Proceedings of the 3rd International Symposium on Multi-Agent Systems, Large Complex Systems, and E-Businesses*, 2002.
10. Paolo Bouquet, Bernardo Magnini, Luciano Serafini, and Stefano Zanobini. A sat-based algorithm for context matching. In *CONTEXT*, pages 66–79, 2003.
11. S. Bressan, C. H. Goh, K. Fynn, M. Jakobisiak, K. Hussein, H. Kon, T. Lee, S. Madnick, T. Pena, J. Qu, A. Shum, and M. Siegel. The Context Interchange Mediator Prototype. In *Proc. SIGMOD 1997*, Tucson, AZ, 1997.
12. A. Brogi, S. Corfini, and R. Popescu. Flexible Matchmaking of Web Services Using DAML-S Ontologies. In *Proc. Forum of the Second Int. Conference on Service Oriented Computing (ICSOC'04)*, New York City, NY, USA, 2004.
13. Diego Calvanese, Elio Damaggio, Giuseppe De Giacomo, Maurizio Lenzerini, and Riccardo Rosati. Semantic data integration in p2p systems. In *DBISP2P*, pages 77–90, 2003.
14. Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, and Riccardo Rosati. Logical foundations of peer-to-peer data integration. In *PODS*, pages 241–251, 2004.
15. F. Casati, M-C. Shan, and D. Georakopoulos. The VLDB Journal: Special Issue on E-Services. *Springer-Verlag Berlin Heidelberg*, 10(1), 2001.
16. Silvana Castano, Alfio Ferrara, and Stefano Montanelli. H-match: an algorithm for dynamically matching ontologies in peer-based systems. In *SWDB*, pages 231–250, 2003.

17. Tiziana Catarci and Maurizio Lenzerini. Representing and Using Interschema Knowledge in Cooperative Information Systems. *Journal of Intelligent and Cooperative Information Systems*, 2(4), 1993.
18. Paolo Ceravolo, Ernesto Damiani, and Marco Viviani. Bottom-up extraction and trust-based refinement of ontology metadata. *IEEE Transactions on Knowledge and Data Engineering*, 19(2):149–163, 2007.
19. The OWL Service Coalition. *OWL-S 1.1 release*, 2004. `http://www.daml.org/services/owl-s/1.1/`.
20. Óscar Corcho. Ontology based document annotation: trends and open research problems. *IJMSO*, 1(1):47–57, 2006.
21. Arturo Crespo and Hector Garcia-Molina. Routing indices for peer-to-peer systems. In *ICDCS*, pages 23–, 2002.
22. Philippe Cudré-Mauroux, Karl Aberer, Alia I. Abdelmoty, Tiziana Catarci, Ernesto Damiani, Arantxa Illaramendi, Mustafa Jarrar, Robert Meersman, Erich J. Neuhold, Christine Parent, Kai-Uwe Sattler, Monica Scannapieco, Stefano Spaccapietra, Peter Spyns, and Guy De Tré. Viewpoints on emergent semantics. *J. Data Semantics VI*, 4090:1–27, 2006.
23. Philippe Cudre-Mauroux, Karl Aberer, and Andras Feher. Probabilistic message passing in peer data management systems. In *ICDE '06: Proceedings of the 22nd International Conference on Data Engineering (ICDE'06)*, page 41, Washington, DC, USA, 2006. IEEE Computer Society.
24. M. Dean, G. Schreiber, S. Bechhofer, F. van Harmelen, J. Hendler, I. Horrocks, D. L. McGuinness, P. F. Patel-Schneider, and L. A. Stein. *OWL Web Ontology Language W3C Recommendation*, 2004. `http://www.w3.org/TR/2004/REC-owl-ref-20040210/`.
25. Hong Hai Do and Erhard Rahm. Coma - a system for flexible combination of schema matching approaches. In *VLDB*, pages 610–621, 2002.
26. AnHai Doan, Pedro Domingos, and Alon Y. Halevy. Reconciling schemas of disparate data sources: A machine-learning approach. In *SIGMOD Conference*, pages 509–520, 2001.
27. AnHai Doan, Jayant Madhavan, Robin Dhamankar, Pedro Domingos, and Alon Y. Halevy. Learning to match ontologies on the semantic web. *VLDB J.*, 12(4):303–319, 2003.
28. Marc Ehrig, Peter Haase, F. van Harmelen, Ronny Siebes, Steffen Staab, Heiner Stuckenschmidt, Rudi Studer, and Christoph Tempich. The swap data and metadata model for semantics-based peer-to-peer systems. In Michael Schillo, Matthias Klusch, Jörg P. Müller, and Huaglory Tianfield, editors, *Proceedings of MATES-2003. First German Conference on Multiagent Technologies*, volume 2831 of *LNAI*, pages 144–155, Erfurt, Germany, SEP 2003. Springer.
29. Marc Ehrig, Christoph Schmitz, Steffen Staab, Julien Tane, and Christoph Tempich. Towards evaluation of peer-to-peer-based distributed information management systems. In *Proceedings of the AAAI Spring Symposium on Agent-Mediated Knowledge Management (AMKM-03)*, Stanford, March 2003.

30. Esteem Team. Emergent Semantics and Cooperation in MultiKnowledge Environments: the ESTEEM Architecture. In *Proc. of the VLDB Int. Workshop on Semantic Data and Service Integration (SDSI 2007)*, Vienna, Austria, 2007. To appear.
31. Ronald Fagin, Phokion G. Kolaitis, Renée J. Miller, and Lucian Popa. Data exchange: Semantics and query answering. In *ICDT*, pages 207–224, 2003.
32. Rino Falcone and Cristiano Castelfranchi. Social Trust: A Cognitive Approach. In *Trust and deception in virtual societies*, pages 55–90. Kluwer Academic Publishers, Norwell, MA, USA, 2001.
33. Weiguo Fan, Hongjun Lu, Stuart Madnick, and David Cheungd. Discovering and Reconciling Value Conflicts for Numerical Data Integration. *Information Systems*, 26(8), 2001.
34. D. Fensel, V.R. Benjamins, E. Motta, and B. Wielinga. UPML: A Framework for Knowledge System Reuse. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI'99)*, pages 16–23, Stockholm, Sweden, 1999.
35. D. Fensel, C. Bussler, Y. Ding, and B. Omelayenko. The Web Service Modeling Framework WSMF. *Electronic Commerce Research and Applications*, 1(2):113–137, 2002.
36. Diego Gambetta. Can we trust trust? In *Trust: Making and Breaking Cooperative Relations*. Basil Blackwell, 1988. Reprinted in electronic edition from Department of Sociology, University of Oxford, chapter 13, pp. 213-237.
37. Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. On reconciling data exchange, data integration, and peer data management. In *PODS*, pages 133–142, 2007.
38. F. Giunchiglia and P. Shvaiko. Semantic Matching. *Knowledge engineering review*, 18(3):265–280, 2003.
39. Tyrone Grandison and Morris Sloman. A Survey of Trust in Internet Applications. *IEEE Communications Surveys and Tutorials*, 3(4), September 2000.
40. Nathan Griffiths. Trust: Challenges and Opportunities. In *AgentLink News*, volume 19, pages 9–11. AgentLink, 2005.
41. Ramanathan V. Guha and Rob McCool. Tap: A semantic web testbed. *J. Web Sem.*, 1(1):81–87, 2003.
42. Alon Y. Halevy, Zachary G. Ives, Dan Suciu, and Igor Tatarinov. Schema mediation in peer data management systems. In *ICDE*, pages 505–, 2003.
43. Katja Hose, Daniel Klan, and Kai-Uwe Sattler. Distributed data summaries for approximate query processing in pdms. In *IDEAS*, pages 37–44, 2006.
44. Matthias Jarke, Maurizio Lenzerini, Yannis Vassiliou, and Panos Vassiliadis, editors. *Fundamentals of Data Warehouses*. Springer Verlag, 1995.
45. James Hendler Jennifer Golbeck, Bijan Parsia. Trust networks on the semantic web. Proceedings of Cooperative Information Agents 2003, August FebruaryJuly-FebruarySeptember 2003.
46. Audun Jøsang, Roslan Ismail, and Colin Boyd. A survey of trust and reputation systems for online service provision. *Decis. Support Syst.*, 43(2):618–644, 2007.

47. Audun Jøsang and S.J. Knapskog. A metric for trusted systems. In *Proceedings of the 21st NIST-NCSC National Information Systems Security Conference*, pages 16–29, 1998.
48. V. Kashyap. Trust, but verify: Emergence, trust, and quality in intelligent systems. *IEEE Intelligent Systems*, 19(5), 2004.
49. U. Keller, H. Lausen, and D. Roman. *Web Service Modeling Ontology (WSMO).* WSMO Working Draft, 2004. http://www.wsmo.org/2004/d2/v02/20040306/.
50. Anastasios Kementsietsidis. Data sharing and querying for peer-to-peer data management systems. In *EDBT Workshops*, pages 177–186, 2004.
51. Anastasios Kementsietsidis, Marcelo Arenas, and Renée J. Miller. Mapping data in peer-to-peer systems: Semantics and algorithmic issues. In *SIGMOD Conference*, pages 325–336, 2003.
52. J. Madhavan, P.A. Bernstein, P. Domingos, and A. Halevy. Representing and Reasoning about Mappings between Domain Models. In *Proc. of the 18th National Conference on Artificial Intelligence and 14th Conference on Innovative Applications of Artificial Intelligence*, Edmonton, Alberta, Canada, 2002.
53. Daniel W. Manchala. Trust metrics, models and protocols for electronic commerce transactions. In *ICDCS '98: Proceedings of the The 18th International Conference on Distributed Computing Systems*, page 312, Washington, DC, USA, 1998. IEEE Computer Society.
54. A. Mawlood-Yunis, M. Weiss, and Nicola Santoro. Issues for robust consensus building in p2p networks. In *OTM Workshops (2)*, pages 1021–1027, 2006.
55. D. Harrison McKnight and Norman L. Chervany. The meanings of trust. Technical Report WP 96-04, University of Minnesota, Carlson School of Management, 1996.
56. George Mihaila, Louiqa Raschid, and Maria Vidal. Using Quality of Data Metadata for Source Selection and Ranking. In *Proc. 3rd International Workshop on the Web and Databases (WebDB'00)*, Dallas, Texas, 2000.
57. E. Motta, J. Domingue, L. Cabral, and M. Gaspari. IRS-II: A Framework and Infrastructure for Semantic Web Services. In *Proc. of the 2nd International Semantic Web Conference (ISWC'03)*, pages 306–318, Sanibel Island, Florida, USA, 2003.
58. E. Motta, J. Domingue, L. Cabral, and M. Gaspari. METEOR-S Web Service Annotation Framework. In *Proc. of the 13th International World Wide Web Conference (WWW'04)*, pages 553–562, New York, NY, USA, 2004.
59. Lik Mui, Mojdeh Mohtashemi, and Ari Halberstadt. Notions of reputation in multi-agents systems: a review. In *AAMAS '02: Proceedings of the first international joint conference on Autonomous agents and multiagent systems*, pages 280–287, New York, NY, USA, 2002. ACM Press.
60. Daniel Olmedilla, Omer F. Rana, Brian Matthews, and Wolfgang Nejdl. Security and trust issues in semantic grids. In Carole Goble, Carl Kesselman, and York Sure, editors, *Semantic Grid: The Convergence of Technologies*, number 05271 in Dagstuhl Seminar Pro-

ceedings. Internationales Begegnungs- und Forschungszentrum fuer Informatik (IBFI), Schloss Dagstuhl, Germany, 2006.
61. Beng Chin Ooi, Yanfeng Shu, and Kian-Lee Tan. Relational data sharing in peer-based data management systems. *SIGMOD Rec.*, 32(3):59–64, 2003.
62. Ewa Orlowska. Kripke semantics for knowledge representation logics. *Studia Logica*, page 25, 1990.
63. Fausto Rabitti and Pasquale Savino. Querying semantic image databases.
64. Erhard Rahm and Philip A. Bernstein. A survey of approaches to automatic schema matching. *VLDB J.*, 10(4):334–350, 2001.
65. Sarvapali D. Ramchurn, Dong Huynh, and Nicholas R. Jennings. Trust in multi-agent systems. *Knowl. Eng. Rev.*, 19(1):1–25, 2004.
66. Lars Rasmusson and Sverker Jansson. Simulated social control for secure internet commerce. In *NSPW '96: Proceedings of the 1996 workshop on New security paradigms*, pages 18–25, New York, NY, USA, 1996. ACM Press.
67. Matthew Richardson, Rakesh Agrawal, and Pedro Domingos. *Trust Management for the Semantic Web*, volume 2870. January 2003.
68. Kai Sattler, Stefan Conrad, and Gunter Saake. Interactive Example-Driven Integration and Reconciliation for Accessing Database Integration. *Information Systems*, 28, 2003.
69. Monica Scannapieco, Antonino Virgillito, Carlo Marchetti, Massimo Mecella, and Roberto Baldoni. The DaQuinCIS Architecture: a Platform for Exchanging and Improving Data Quality in Cooperative Information Systems. *Information Systems*, 29(7):551–582, 2004.
70. P. Shvaiko and J. Euzenat. A Survey of Schema-based Matching Approaches. *Journal on Data Semantics (JoDS)*, IV:146–171, 2005.
71. Steffen Staab. Emergent semantics. *IEEE Intelligent Systems*, 17(1):78–86, 2002.
72. Steffen Staab, Bharat Bhargava, Leszek Lilien, Arnon Rosenthal, Marianne Winslett, Morris Sloman, Tharam S. Dillon, Elizabeth Chang, Farookh Khadeer Hussain, Wolfgang Nejdl, Daniel Olmedilla, and Vipul Kashyap. The pudding of trust. *IEEE Intelligent Systems*, 19(5):74–88, 2004.
73. Richard Y. Wang, Henry B. Kon, and Stuart E. Madnick. Data Quality Requirements: Analysis and Modeling. In *Proc. ICDE '93*, Vienna, Austria, 1993.
74. Glynn Winskel. *The formal semantics of programming languages: an introduction.* MIT Press, Cambridge, MA, USA, 1993.
75. Ling L. Yan and T. Ozsu. Conflict Tolerant Queries in AURORA. In *Proc. CoopIS'99*, Edinburgh, UK, 1999.
76. L. Zeng, B. Benatallah, M. Dumas, J. Kalagnanam, and H. Chang. QoS-Aware Middleware for Web Services Composition. *IEEE Transactions on Software Engineering*, 30(5):311–327, 2004.