Open Ontologies – The Need for Modeling Heterogeneous Knowledge

Tina Froehner[1] Matthias Nickles[2] Gerhard Weiß[2]

[1]*Research Center Knowledge Management (RCKM), University of Applied Sciences Cologne D-50678 Koeln, Germany, tina.froehner@fh-koeln.de*

[2]*AI/Cognition Group, Computer Science Department, Technical University Munich D-85748 Garching, Germany, {nickles,weissg}@tiki.informatik.tu-muenchen.de*

Abstract

In the context of open environments like the Semantic Web, knowledge-based applications with autonomous knowledge sources have recently gained increased interest. For such sources, neither the consistency nor the reliability or the definiteness of the input knowledge can be ensured. Whereas there already exist promising approaches to the filtering, integration and homogenization of inconsistent and dynamic knowledge, the modeling of knowledge heterogeneity and dynamics itself has been largely neglected so far. In this work, we propose Open Ontologies as an approach to this issue. Open Ontologies are based on a distinct communication-oriented paradigm as they emerge from and evolve with communication processes, and allow the modeling and processing of semantically heterogeneous conceptual knowledge by means of its reification according to its social (i.e. communicational) meaning and relevance, and the probabilistic weighting of inconsistent knowledge facets.

Keywords: Formal Ontologies, Semantic Web, Agent Communication, Knowledge Management, Knowledge Integration

1. Introduction

An increasing number of knowledge-based applications gain their informational input from a set of autonomous knowledge sources. Autonomous knowledge sources are entities like software agents, human agent clients or humans themselves operating deliberatively under self-control towards their private goals in a (from an external point of view) more or less unforeseeable manner, which might be malevolent, insincere and not trustworthy. In open application environments such as the Semantic Web, the autonomy of information sources is not only unpreventable, but also enables increased application flexibility and robustness. Despite the indisputable advantages of knowledge source autonomy, there still remain fundamental problems in regard to the handling of data coming from open environment which have not yet been sufficiently addressed by traditional approaches to ontology modeling and acquisition yet, as i) they seldom handle changes of the meaning of concepts, ii) they seldom consider conceptual knowledge as being contextualized with intentions, processes and effects from the "outside world", iii) they often pursue a "one ontology fits all" policy without an adequate consideration of user needs variety, and iv) they have no concept for the treatment of semantic heterogeneity which is sufficient in our opinion. Whereas approaches like *Emergent Semantics* [1], *Dynamic Ontologies* [2] and semantical *ontology merging* [e.g. 3] have caused significant improvements regarding these issues, semantical inconsistencies are almost always still taken for something which either should be avoided, or should be homogenized using, e.g. clustering techniques, or should be filtered out (e.g., using criteria like (dis-)trust or source reputation [8]), and they are almost always treated as unintended, incomputable situations "without a cause". In demarcation from such views, it should be recognized, that semantical inconsistencies are not just unfavorable states, but that they are in real-world environments often unpreventable due to stable belief or goal conflicts [5] of knowledge sources, that they can even provide the knowledge user with valuable meta-information about the intentions, goals and social relations among the knowledge sources, and, if they have been made explicit and visible, that they can be prerequisites for a subsequent conflict resolution. In general, mechanisms for knowledge integration, e.g. within the field of knowledge management, can only be a preliminary decision about the reasonable modeling of communicated knowledge artifacts, because within a heterogeneous group of knowledge users, in the end each user can only decide for himself about the relevance and correctness of the given information, which provides a strong arguments for the conservation of heterogeneous knowledge while integrating.

Since the Semantic Web can be seen as typical for open environments most of the issues arising from knowledge source autonomy can be found there, and due to the extend and the dynamics of the Semantic Web they are expected to appear in their most severe shaping. The same problems arise in knowledge management, when creating and sharing knowledge within e.g. large organizational enterprise knowledge bases, or between organizations (e.g. virtual organizations, outsourcing and offshore), and in general in the context of the automation of business data retrieval, integrating and analyzing [15]. Nevertheless, for instance in large organizational enterprise knowledge bases in principle similar problems can arise. Although the Semantic Web effort addresses the problem of missing machine-understandability of web site descriptions, it currently focuses primarily on the specification of languages and tools for the representation of consistent semantics and ontologies, not on inherently social processes of information gathering and describing itself, and it is just beginning to take into consideration phenomena like the social impact of resource descriptions [6], conflicting opinions, information biasing by competing commercial interests, and inconsistent or intentionally incorrect information. So, the fact that probably the most important aspect of the Semantic Web is the explicit or implicit communication among a possibly very large number of autonomous actors is still largely neglected in ontology research. Bringing information (e.g. via websites or -services) into the web is in fact a social act, and the relationship between informational artifacts on the web is communicational (i.e. *specifying, agreeing, contradicting*…). This can of course lead to intentional and unavoidable inconsistencies of ontological concepts (e.g. company interests versus customer interests or various conceptualizations due to differences in culture). If these are filtered out or homogenized to early, important information for the user or the application might be lost.

In response to the mentioned issues, we aim at a combination of recent advances in distributed artificial intelligence regarding the modeling of agent communication and the research on knowledge management for open environments, we propose *Open Ontologies* that are derived from the communication of multiple knowledge sources and users, and maintain semantical heterogeneity and social structures.

The reminder of this paper is structured as follows: The next section outlines the basic properties of Open Ontologies, and the principles of their generation and representation, section 3 shows how the "truth layers" of the Semantic Web would change with our concept applied, and section 4 closes with an outlook on open research issues.

2. Open Ontologies

Formal ontologies are traditionally defined as agreed formal conceptualizations of certain domains, which serve as common ground for distributed tasks like knowledge exchange and modeling. As we have seen, this understanding leads to difficulties if the informational input the ontology is build from is likely to be intentionally inconsistent, and there either does not yet exist enough meta-knowledge like trust to identify and filter out "inappropriate" or "wrong" data *a priory*, or there does not even exist a concept of global inappropriateness or correctness at all. On the other hand, consistent and agreed ontologies are doubtless an inevitable prerequisite for efficient knowledge creation, representation and exchange, whereby we consider implicite and emerged ontologies and schemata (e.g. in the context of semi-structured data modeling) to be such ontologies too. Of course, in principal ontological heterogeneity can be overcome by means of techniques like the consistent renaming of inconsistent concepts. However, such solutions often generate redundancy instead of an informational benefit for the knowledge users, if the communicational relations among previously inconsistent assertions are lost. Open Ontologies aim at the solution for this dilemma by embedding conceptual knowledge facets gained from a heterogeneous set of self-interested autonomous knowledge sources (e.g. information agents or humans) within contextual information about their communicational (i.e. social) origin and impact, their probability and assertive weight, and relationships (e.g., contradiction, approval, revision or specification) to other communicated facets. Such meta-information can be gained from the evaluation of formal communication, but could also be derived from, e.g., structured, semi-structured or natural language documents. Doing so, in Open Ontologies information as it can be found as first-order concept descriptions in conventional Ontologies, is lifted to the *social level* and thus to a level where the sources and the users of the ontology are likely to achieve an agreement with the social assessments of possibly inconsistent and uncertain facts. The agreement or disagreement with certain assessed facts, based on the meta-information the Open Ontology provides is then a subsequent task based on rich social knowledge instead of binary decisions like to trust or not to trust particular agents. Open Ontologies are thus *formal ontologies which receive their content from the communication of multiple autonomous sources and users, and provide a dynamic representation of socially annotated heterogeneous knowledge.*

In the philosophical tradition of Pragmatism [7] and the works of Wittgenstein on linguistic meaning [9], communication is here not so much to be understood as the exchange of symbols with a fixed meaning, but the other way round as a means to generate supra-individual meaning from interrelated interactions among black- or gray-box agents (i.e., agents with more or less unknown internal states, cognition and goals). The practical consequences arising from this are that Open Ontologies are necessarily *dynamic ontologies* which need to be continuously adapted to new information, and that the processes of creation, understanding and usage of ontological information are *integral aspects* of Open Ontologies, i.e., Open Ontologies are not just evolving from and influencing communication processes, but these processes (precisely: suitable representations of certain process structures) are part of the ontologies themselves.

In addition, communication among multiple agents likely requires mechanisms for the *generalization* of emergent meaning, since otherwise the complexity of the ontology would grow too large due to the sheer number of individual knowledge contributions. Furthermore, generalization is also a way to make Open Ontologies look like homogeneous ontologies if necessary, because at its highest level, generalization

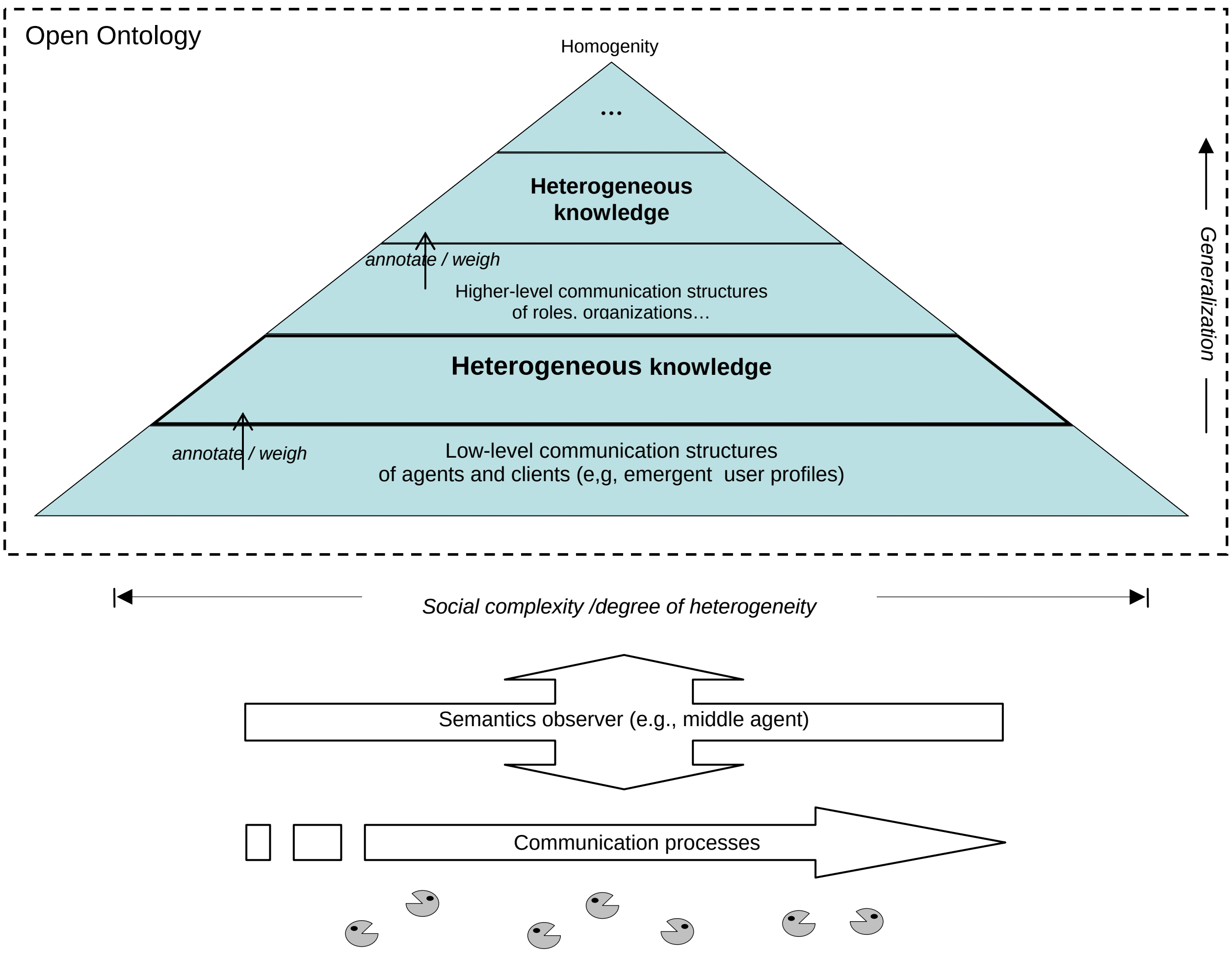

Figure 1. Emergence of Open Ontologies

causes semantical homogenization among contradicting knowledge sources. Summing it up, Open Ontologies have the following main characteristics:

- **Openness:** No assumptions are made regarding the architecture, number, benevolence, trustworthiness, relevance, informedness and cooperativeness of its sources. Nevertheless, mechanisms for the establishment of source identification like digital signatures may have to be applied.

- **Dynamical derivation from communication:** Open Ontologies are emergent from and evolving with ongoing communication of information agents (knowledge sources and knowledge users) to assert (deny, specify…) information and to express and specify informational needs.

- **Maintenance, explicitness and social annotation of semantical heterogeneity:** Open Ontologies maintain semantical inconsistencies arising from knowledge-source contradictions and conflicts (in leveled degrees of complexity, see below), and contain annotations of conceptual knowledge with meta-information about its meaning (especially their weight) within the course of communication.

- **Multiple, probabilistically modeled levels of social generalization:** They allow multiple levels of generalization of social concepts (e.g. groups and roles), weighting the degree of assumed social acceptance of the annotated ontological information.

In the following, we will sketch the framework for the generation and representation of Open Ontologies, for lack of space without providing technical or formal details. An application of some of the concepts behind Open Ontologies to the multiagent-based rating of RDF statements can be found in [11, 10]. The emergence of semantics from the empirical observation of agent communication is described in detail in [12].

2.1. Representation and Acquisition

Typically, Open Ontologies have a data ← meta-data structure, containing as first-order objects knowledge facets that have the form '*1st-level knowledge ← 2nd-level knowledge*', where 1st-level knowledge partially describes a domain concept in the same way as within usual ontologies, but probably in an inconsistent way regarding other 1st-level knowledge in the same ontology (please see [4] for details). Since Open Ontologies are primarily an abstract meta-concept build upon conventional approaches for the representation of conceptual knowledge, we do not constrain or specify the sort of concrete entities that are to be "wrapped" within an Open Ontology or at the content level of agent messages, like FOL statements, classes or frames in this work. For the same reason, we do also not make any assumptions relating to ontology domains or concrete areas of application here. In contrast to 1st-level knowledge, 2nd-level knowledge (also called *social knowledge*) depicts the social context of 1st-level knowledge, the latter taken as generated from a communication act of an autonomous source of knowledge. This kind of annotation of 1st-level knowledge with 2nd-level knowledge we call *social reification* [4]. The most trivial kind of social reification is *quoting* (e.g., 'Sue says: "…"', e.g., expressed by means of reified RDF statements), but in general, all kind of information which describes how and to what effect certain data is produced within a process of communication can be informally understood as 2nd-level knowledge (and, of course, we can apply social reification recursively, i.e. annotate 2nd-level knowledge with 3rd-level knowledge as in 'Sue says: 'Tom says: "…"'' and so on). The most elementary forms of such social meta-data are considered agent speech act types like *assertion, denial* or *query*, inducing relations among single communication like 'Sue contradicts Tom's statement saying "…"' and rich 2nd-level knowledge types such as knowledge source and user profiles and even complex social systems like organizations [13]. In our empirical communication model [12] these symbolic acts gain their meaning from their expected effect on the subsequent trajectory of communications, which can be learned empirically from past interactions (although we recommend empirical semantics to disregard mentalistical details which are unknown for autonomous agents and allow for the handling of uncertain meanings, the usage of such a semantics is not required to define an Open Ontology). Because meaning is contextualized by the situation (history) of the respective act occurrence, in general 2nd-level knowledge describes communication processes (this applies even to simple quotations: In Sue says: "…", "Sue" is in fact just an abbreviation for the pragmatic impact utterances from Sue are expected to have. As mentioned earlier, Open Ontologies require the generalization of meaning in order to reduce complexity. Generalization in this sense has two dimensions: i) the merging of 2nd-level knowledge, and

ii) the subsequent merging of related 1st-level knowledge facets. Typically, i) comprises the merging of similar social processes to *interactions patterns*, and the combination of multiple similar behaving agents to social *groups* or *roles*. After applying such generalization rules to 2nd-level knowledge, the annotated 1st-level knowledge needs to be merged accordingly. If, for example, multiple agents forming a single social group make inconsistent assertions, within the Open Ontology each of these assertions obtains a probabilistic weight expressing the degree of expected approval this assertions gets from the role or group as a whole (calculated, e.g., from the frequency this assertion has been uttered by different agents within this role or group) [11, 10].

We propose the usefulness of a co-presence of multiple levels of generalization, tailored to the desired levels of heterogeneity of the Open Ontology.

Open Ontologies are continuously learned by an observer of knowledge source and knowledge user communications. The technical requirements for this learning process are i) *information agents* able to communicate 1st-level knowledge facets like "c_1 is-a c_2" or "c_1 has-a c_2" at the content level of their communication language (since Open Ontologies do not require agent cooperativeness, performatives used for, e.g. negotiation are not required, although they would make up a useful extension), ii) a facility for the acquisition of Open Ontologies from the observation of agent communication, e.g., a dedicated middle agent within the infrastructures of the respective application, or framework called a *semantics observer*, iii) a facility for the low-level storage and querying of persistent ontological data (e.g., a database management system), and optionally iv) a facility for the social reasoning upon the 2nd-order knowledge within the Open Ontology (to deduce new facts like "Sue is likely to contradict or specify Toms information", but also to derive trust relationships among the participants).

The acquisition of Open Ontologies (cf. figure 1) comprises the following main tasks, which are to be performed in a loop as a continuous, incremental learning process for the whole period of agent communication:

1) Observation and filtering of agent communication (e.g., due to restrictions to a certain concept domain)

2) Derivation and/or adaptation of 1st-level and 2nd-level knowledge according to the respective semantical model (e.g. empirically)

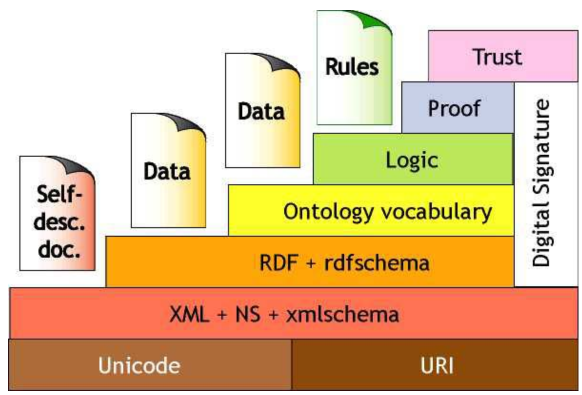

Figure 2. Semantic Web layers

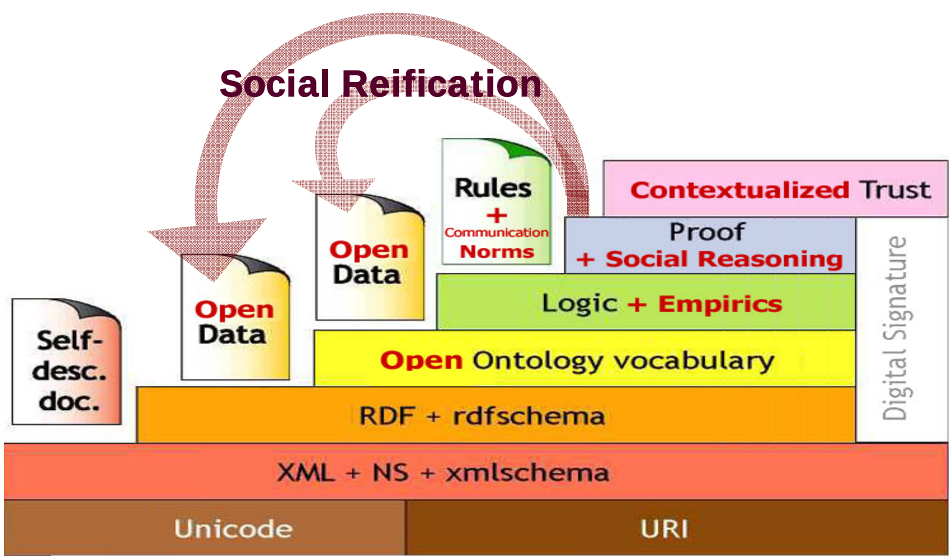

Figure 3. Social reification and reasoning for the Semantic Web

3) Stochastic generalization of 2nd-level knowledge (multiple levels, if necessary)

4) Social reification and generalization of 1st-level knowledge (multiple levels, if necessary)

5) Alignment with normative 1st-level knowledge (like an obligatory top-level ontology) or given 2nd-level knowledge (social, i.e., communicational norms), if necessary

3. Open Ontologies for the Semantic Web

Figure 2 shows the semantical levels proposed by Tim Berners-Lee for the structure of the forthcoming Semantic Web [14]. We recommend for some aspects of this concept an extension or specification to provide solutions for the issues mentioned earlier in this work (figure 3). In particular, it appears to be inevitable to us to provide formalisms and calculi that explicitly consider semantically heterogeneous meta-data like resource descriptions and ontologies created from the contributions of multiple sources that compete for the assertion of their individual “truths” and interests. Of course, the Semantic Web would already be implicitly some sort of Open Ontology, but for a broad acceptance and to provide value to its users, we strongly suppose that communicational relationships among closed “islands” of knowledge like contradiction or agreement need to be made explicit formally and technically. In this regard, the empirical derivation and stochastic modeling of “open” meta-data also seems inevitable if the set of knowledge sources is either very large, or fluctuates, or generates indefinite information. These mechanisms are not meant to be replacements for the usage of first-order predicate logic for Web reasoning, but instead as a completion which could be introduced gradually. E.g., the Resource Description Framework RDF(S) already has low-level reification capabilities, which could be used for elementary social reifications as described in [11,10], but in general this would require an appropriate specification of this kind of usage and technical and theoretical support for its realization.

4. Conclusion

There is an obvious and rapidly growing need for knowledge-based systems capable of running in open environments with autonomous knowledge sources and users, given the increasing interoperability and interconnectivity among computing platforms. Taking the key properties of such environments like information agent autonomy and the emergence of meaning from interaction seriously concerning the development of ontologies is a great challenge. On the one hand, ontologies should provide a stable ground for communication and subsequent knowledge modeling, on the other hand, in open environments concept descriptions tend to be semantically inconsistent, they emerge from competing beliefs and goals, and a priori there is no such thing as a commonly agreed “truth” (in the “real world”, not even a trend towards such a thing can be assumed). To cope with each of these two contradictory aspects must be a core concern of the communication-oriented paradigm of knowledge modeling and management, and is the basic motivation underlying the work described here. For this purpose, we have proposed Open Ontologies as a fundamental step towards the modeling and representation of semantical knowledge heterogeneity. To explore and to work out such a new perspective constitutes a long-term scientific and practical endeavor of considerable complexity. Therefore, it is not surprising that this introductory work does not answer all relevant issues and leaves room for theoretical and practical specification.

References

[1] A. Maedche, F. Nack, S. Santini, S. Staab, L. Steels. Emergent Semantics. IEEE Intelligent Systems, Trends & Controversies, 17(2), 2002.

[2] J. Heflin, J. A. Hendler. Dynamic Ontologies on the Web. Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence, p. 443 - 449, 2000.

[3] V.A.M. Tamma. An Ontology Model Supporting Multiple Ontologies for Knowledge Sharing. PhD Thesis, University of Liverpool, 2002.

[4] M. Nickles, T. Froehner. Social Reification for the Semantic Web. Research Report FKI-24x-04, AI/Cognition Group, Department of Informatics, Technical University Munich, 2004.

[5] C. Castelfranchi. Conflicts Ontology. In R. Dieng, H.J. Mueller, Conflicts in Artificial Intelligence. Springer, 2000.

[6]http://www.w3.org/2001/sw/meetings/tech-200303/social-meaning/

[7] G. H. Mead. Mind, Self, and Society. University of Chicago Press, Chicago, IL, 1934.

[8] J. Golbeck, B. Parsia, J. Hendler. Trust Networks on the Semantic Web.Proceedings of Cooperative Intelligent Agents, 2003.

[9] L. Wittgenstein. Tractatus Logico-Philosophicus. Routledge, Cambridge, Mass., 2001.

[10] M. Nickles, Towards a Multiagent System for Competitive Website Ratings. Research Report FKI-243-01, Technical University of Munich, 2001.

[11] M. Nickles, G. Weiss. A framework for the social description of resources in open environments. Proceedings of the Seventh International Workshop on Cooperative Information Agents (CIA, pp. 206-221). Lecture Notes in Computer Science, Volume 2782. Springer, 2003.
[12] M. Nickles, M. Rovatsos, G. Weiss. Empirical-Rational Semantics of Agent Communication. Proceedings of the Third International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'04), New York City, 2004. To appear.
[13] N. Luhmann. Social Systems. Stanford University Press, Palo Alto, CA, 1995.
[14] T. Berners-Lee, J. Hendler, O. Lassila. The Semantic Web. Scientific American 184(5), 2001.
[15] H. Pham, Y. Ye, V. Nguyen. Autonomous Mapping of E-business Demands and Supplies via Invisible Internet Agents, International Journal of E-commerce Research, Kluwer Publisher, Volume 3:3-4, pp.365-395, 2003.