Open Ontologies and Open Knowledge Bases for Modeling the Social Layer of the Semantic Web

Matthias Nickles[‡] and **Tina Froehner**[⋆] and **Gerhard Weiss**[‡]
[‡] AI/Cognition Group, Technical University Munich (TUM), {nickles,weissg}@cs.tum.edu,
[⋆] Research Center Knowledge Management (RCKM),
University of Applied Sciences Cologne, tina.froehner@fh-koeln.de

Abstract

We propose *Open Ontologies* and *Open Knowledge Bases* as means for the acquisition, evolution and representation of socially (but not necessarily cooperatively) generated semantically heterogeneous knowledge for open environments like the Semantic Web. In contrast to traditional approaches to knowledge assembling, knowledge sharing and ontology integration, which emphasize semantical consistency and the consensus finding among the participants, Open Ontologies and Open Knowledge Bases derive knowledge from possibly inconsistent, indefinite or conflictive communication processes. They explicitly model semantical heterogeneity arising from source conflicts in multiple levels of complexity reduction, and allow the probabilistic weighting of inconsistent knowledge.

1 The Need for Modeling Social Knowledge Heterogeneity

Decisive for the success of the Semantic Web is the provision of formalisms, mechanisms and tools for the acquisition, modeling and handling of heterogeneous knowledge coming from a very large number of distributed, autonomous and potentially conflicting sources. The Semantic Web needs to cope with the fact that the meaning of information on the web can never be determined for sure in general, might change, and might be constituted from the possibly conflicting opinions of large sets of knowledge sources. In contrast, most of the current efforts in order to build the Semantic Web concentrate on the specification of languages and tools for the modeling of semantically consistent knowledge, and research is just beginning to take into consideration phenomena like the social (i.e. communicative) impact of resource descriptions, conflicting opinions, information biased by competing commercial interests, and inconsistent or intentionally incorrect information. Bringing information into the web is in fact a social act, and the relationship between informational artifacts on the web is communicative (i.e. specifying, agreeing, contradicting, querying...). This can of course produce 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 ignored, or filtered out, or homogenized to early (e.g. by applying *trust* relationships or recommendations), important information about the "social landscape" of knowledge might be lost.

These issues have in common that they rise mainly from the *autonomy* and *pro-activity* of knowledge sources and users, being black- or gray-box actors with more or less opaque goals they pursue asserting or forming their individual world views. The way such autonomous entities (especially *information agents*) exchange information is *communication*. Web knowledge can be considered as communicative among such autonomous entities, because it is generated in order to influence its recipients, with often unknown intentionality, reliability and user reception. This is even true if knowledge is communicated indirectly and asynchronously using e.g. web pages or databases instead of information agents. Web knowledge is also contextualized with other web knowledge, and it can be agreed and denied by other knowledge on the web.

Whereas approaches like *Emergent Semantics* [Maedche *et al.*, 2002], *Dynamic Ontologies* [Heflin and Hendler, 2000] and semantical ontology merging and alignment have caused significant improvements regarding some of these problems, semantical inconsistencies due to conflicting knowledge sources are almost always still taken for something which either should be avoided, or should be homogenized, or should be filtered out, e.g., using criteria like (dis-)trust and source reputation [Golbeck *et al.*, 2003]. 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 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 subsequent conflict resolution and recommendations.

We propose *Open Ontologies* and *Open Knowledge Bases* (OO/OKB for short, if both terms are applicable) as a fundamental approach to the *social* acquisition and modeling of knowledge for open environments like the Semantic Web, based on a combination of recent advances in distributed artificial intelligence regarding the derivation of agent communication semantics and research on knowledge management for open environments. Open Ontologies (focussing on concept knowledge) and Open Knowledge Bases (focussing on concept instances) are derived

from the communication of multiple autonomous knowledge sources and users, and maintain semantical heterogeneity and social knowledge structures in multiple levels of complexity reduction.

2 Open Ontologies and Open Knowledge Bases

Shared knowledge like in ontologies is traditionally defined as an agreed description of certain domains which serve as common ground for tasks like communication and user information. This understanding leads to difficulties if the informational input the knowledge is gained from is likely to be intentionally inconsistent, and there either does not yet exist enough additional information like trust or recommendations to identify, rank or filter out "inappropriate" or "wrong" information a priory, or there does not even exist a concept of global inappropriateness or correctness at all.
Open Ontologies [Froehner *et al.*, 2004] and Open Knowledge Bases [Nickles *et al.*, 2004a] aim at a solution for this dilemma by embedding (possibly inconsistent) knowledge contributions gained from a heterogeneous set of self-interested *autonomous* knowledge sources within (partially optional and probabilistic) meta-information about their *social* meaning, especially about their sources (e.g. web resources or agents), their assertive weights (how "strong" is the respective opinion, and how likely can it be ascribed to a certain source?), their social impact and dissemination, their discursive relations to other communicated knowledge (e.g., denial, approval, revision or specification), and, if possible, the likely intentions of their assertions. The derivation of, e.g., trust relationships from such social structures is then an optional, subsequent task.
Both the contributions and their accompanying social meanings can e.g. emerge from communication processes using formal agent communication languages (for a stochastical approach to the derivation of social structures from agent communication see [Nickles *et al.*, 2004c]), but might also be derived from structured, semi-structured or natural language documents. Doing so, information as it can be found as first-order statements in traditional ontology and knowledge bases, is *lifted* to the *social level*, a process called *Social Reification* [Nickles *et al.*, 2004a], as it is based on reification facilities that can be found e.g. in the RDF (*Quoting* is considered to be a very simply variant of this principle). Therefore, first-order objects within OO/OKB have a '1st-level knowledge ← 2nd-level knowledge' form, whereby 1st-level knowledge describes a domain in the same way as within usual ontologies/knowledge bases, but probably in an inconsistent way regarding other 1st-level knowledge within the same OO/OKB. In contrast, 2nd-level knowledge describes the social meaning of the respective annotated 1st-level knowledge. This concept is loosely related to *context logic*, but does not aim for the provision of local truth-contexts. Rather, 2nd-level knowledge states the sound social meaning for true as well as false 1st-level statements. OO/OKB are thus *dynamic, shared information bases which receive their emergent content from the direct or indirect communication of multiple autonomous knowledge sources and users, and provide a dynamic representation of socially annotated, semantically heterogeneous knowledge.*

In addition, OO/OKB provide mechanisms for their leveled generalization, since otherwise their complexity would possibly grow too large due to the sheer number of heterogeneous individual knowledge contributions and the richness of social structures. Generalization is also a way to make a certain OO/OKB appear like a homogeneous ontology or knowledge base if necessary, because at its highest level, generalization causes semantical homogenization of contradicting knowledge. E.g., single information agents can be generalized as agent *roles*. The task-specific knowledge associated with a certain role is assumably more general, sound and abstract than the complete knowledge associated with two or more concrete agents (which are able to impersonate this role temporarily). In all, OO/OKB have the following characteristics:
Openness: As few as possible assumptions are made regarding the benevolence, trustworthiness, relevance, informedness and cooperativeness of its sources.
Dynamical derivation from communication: OO/OKB are emergent from and evolving during ongoing (possibly asynchronous or implicit) communication of autonomous knowledge sources and users which assert (deny, specify, query...) subjective information (respectively information needs).
Explicitness and social annotation of heterogeneity: OO/OKB *maintain* semantical inconsistencies arising from contradictions and conflicts, and contain annotations of (conceptual or instance) knowledge with 2nd-level meta-information about its social meaning within communication processes.
Multiple levels of generalization: They allow multiple levels of generalization, defining the degree of inconsistency and social meaning preciseness.
To work out such a new perspective constitutes a long-term scientific and practical endeavor of considerable complexity, leaving room for further specifications and variations.

References

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

[Heflin and Hendler, 2000] J. Heflin, J. A. Hendler. Dynamic Ontologies on the Web. In *Procs. of the 7th National Conference on Artificial Intelligence*, 2000.

[Golbeck *et al.*, 2003] J. Golbeck, B. Parsia, J. Hendler. Trust Networks on the Semantic Web. In *Procs. of Cooperative Intelligent Agents*, 2003.

[Nickles *et al.*, 2004a] M. Nickles, T. Froehner. Social Reification for the Semantic Web. *Research Report FKI-248-04*, Technical University Munich, 2004. To appear.

[Froehner *et al.*, 2004] T. Froehner, M. Nickles, G. Weiss. Towards Modeling the Social Layer of Emergent Knowledge Using Open Ontologies. In *Procs. of the ECAI 2004 Workshop on Agent-Mediated Knowledge Management (AMKM'04)*, 2004.

[Nickles *et al.*, 2004c] M. Nickles, M. Rovatsos, G. Weiss. Empirical-Rational Semantics of Agent Communication. In *Procs. of the Third International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'04)*, 2004.