Summary of the Results of the Break-out Session "Social Issues around the Semantic Web"

Karl Aberer, Klemens Böhm, Fred Fonseca

Karl Aberer
Distributed Information Systems Laboratory (LSIR)
EPFL
1015 Lausanne
Switzerland

Klemens Böhm
Institute of Programming Structures and Data Organisation
University of Karlsruhe
Germany

Fred Fonseca
The Pennsylvania State University
307E Info Sciences and Technology Building
University Park, PA
U.S.A.

Abstract. As part of the Dagstuhl Workshop on the Semantic Web, a break-out session focused on discussing social issues around the Semantic Web. This article is a concise summary of the main issues discussed, the controversies that have arisen, and of the open research questions that need to be addressed.

1 Introduction

The break-out session 'Social Issues around the Semantic Web' took place in Dagstuhl on September 23rd. 2004. The discussion in the session focused on two main themes. The first one is *ontology construction and maintenance*, and the second one is *Peer-to-Peer communities (P2P communities)* and associated applications.

Ontologies are fundamental for the Semantic Web. Nevertheless, many questions regarding their construction process are still open. Furthermore, even if we have managed to build an ontology, the question how to maintain them remains. It is however important since meanings are always changing. Regarding this evolvement, another important issue is how to keep the different versions of ontologies so that past knowledge (or the history of its evolution) is not lost.

Dagstuhl Seminar Proceedings 04391
Semantic Interoperability and Integration
http://drops.dagstuhl.de/opus/volltexte/2005/33

P2P communities are closely related to the Semantic Web as well. They are based on the notion of resource sharing in communities. In order to share information, communities have to establish agreements on information representation. In other words, peer-to-peer systems require ontologies. One kind of resource shared in peer-to-peer systems is knowledge on information representation. This shared knowledge can be used to establish agreements on information representation or to map between different representations. Thus, peer-to-peer interactions can lead to the emergence of ontologies.

In the following, we discuss the potential and challenges of peer-to-peer systems to foster semantic interoperability.

2 Ontologies

Before information sharing takes place within a community, explicit formalizations of the mental concepts that users have about the real world are required. Furthermore, communities must group these concepts such that they represent the basic agreements that exist within each community. Once these mental models have been formalized explicitly, one must create mechanisms for generalization (e.g., a specific type of lake becomes a body of water) and for adding further specification (e.g., the concept of a body of water becomes a specific lake). Humans perform such operations in their minds all the time. Their formalization is necessary to make computer implementations of these operations available. The explicit formalization of the mental models of a certain community is called ontology. The basic description of the real things in the world, the description of what would be the truth, is called Ontology (with an uppercase O). The result of making the agreement within communities explicit is an ontology (with a lower-case o), borrowing from the terminology of the Artificial Intelligence community. Thus, there is only one Ontology, but many ontologies.

Now that there is a way to describe the content of internet resources, the next question is how to share meaning. For efficient information sharing that delivers the kind of data that the users are expecting, it is necessary to have an agreement on the meaning of the data. In a broader scope, it is necessary to reach an agreement on the meaning of the entities that represent the content of web-information resources. These entities are parts of a mental model that represents concepts of the real world. A concept such as ‘body of water’ comprises a definition and the mental image that users have of it. But what kinds of agreement can be reached among users? The question whether it is possible to reach such an agreement among all users regarding the basic entities of the world is a current subject of discussion among researchers. We can see this issue from two different perspectives. From the first one, there is one Ontology, and we can reach consensus on it by refining concepts step by step, over time. The second perspective does not accept this one Ontology. From this perspective, we must live with incompatible views of reality and should try to map concepts from one ontology into other ones whenever possible. – A solution that can reach out to both perspectives supports that small communities make small agreements. These agreements are ex-

panded later, in order to reach larger communities. When this larger agreement occurs, some of the original meaning or at least some level of detail is lost. For instance, within a community of biology scholars, a specific body of water can serve as the habitat of a specific species. Therefore, it can have a specific concept or name referring to it. Nonetheless, it is still a body of water. When a biologist is working at a more general level, it is perceived as such. At this higher level, this real-world entity – body of water – is more likely to find a match with the same concept in another community. So the biologist and a member of the other community can exchange information about bodies of water. The information will be more general, compared to the situation where the body of water is seen as the habitat of a specific species.

3 Peer-to-Peer Systems

Peer-to-peer systems build on a social paradigm. Therefore they seem to be predestined to deal with social phenomena. The problem of agreeing on meaning by using ontologies is a primary example of a social phenomenon. Meaning is not an absolute truth. It rather shifts over time. Further, meaning can be approximate. Depending on the context, an approximate understanding of the meaning of a data object may suffice to facilitate reasonable decisions. Peer-to-peer systems with their fluid structure seem to mimick the characteristics of agreement processes. This should allow to establish semantic agreements well.

As in any social process, trust is the key element for successful interactions. This concerns the issue of interpretation of data as well. Signals received through possibly many intermediaries in large-scale networks need to be weighted according to the trustworthiness of the sources. Semantics is not only about agreement, but requires the ability to disagree as well, when information sources are not deemed reliable. Hence, establishing trusted relationships in peer-to-peer networks has become an important research area. Establishing trust in a peer-to-peer network may be seen as a highly recursive process. Everybody assesses the trustworthiness of other participants and is assessed by other participants at the same time. The simple PageRank model used by Google illustrates this well.

Even with mechanisms in place to form agreements in a trusted manner, useful interactions do not occur without participants being motivated to participate in the process. Experiences from solving semantic interoperability problems in organizational contexts based on human interactions show that incentives are key. These incentives depend on the organizational, business or social context of the participants, but also on the legal framework in many situations. Only by observing the context, one can understand why certain types of peer-to-peer applications have succeeded, while others have not.

One of the crucial incentives, or better dis-incentives, is a sufficient level of privacy. Participants in peer-to-peer information exchange face a fundamental dilemma: on the one hand, they would like to access shared and aggregated data from many partici-

pants. On the other hand, they would like to prevent abuse of personal information. Clearly, the current privacy-protection schemes, such as P3P, are too limited and too inflexible in order to cope with realistic requirements in many applications. Peer-to-peer architectures in turn be a possible solution to some privacy issues that occur in today's centralized architectures. In Web search, to give an example, peer-to-peer architectures could significantly improve access to the hidden Web. Such a solution would allow participants to protect their data at the level required.

Considering the various social aspects of peer-to-peer computing related to establish a "semantic landscape", one might ask what social scientists and economists could contribute based on their specific experience. We imagine two possible paths: The first one is to learn from models developed in social sciences to design systems based on these models. We might call this *socially-inspired computing*. The second one is to study the role of humans as part of the overall system and to provide insight in the various ways they might influence it. Similar statements could be made for economists. Applying economic models in computer science has received significant attention in the past, in particular when it comes to complex resource-allocation problems. An interesting direction in social science is to validate that the Web does not only improve existing processes in quantitative terms, but provides new qualitative perspectives as well. For instance, it opens completely new perspectives for handicapped persons regarding social interactions. Studying such effects in the context of peer-to-peer computing is challenging.

A final question is how the Semantic Web and peer-to-peer systems might help to improve social relationships. In order to to make these technologies fruitful from a social perspective as well, people have to feel comfortable in their use. To this end, establishing mechanisms that allow for social control seem to be important. For example, think of Open Directory. The fact that people owned different projects was problematic and prevented people from participating. When designing ontologies, few individuals tend to be in the way of progress. So far, social control, if people meet in person, is more effective. Questions such as storing the history of decision-making processes reliably seem to be essential to improve social relationships and to make the technologies envisioned here a success from the practical perspective as well.