

Data Integration in a Networked World

Karl Aberer

School of Computer and Communication Sciences
Ecole Polytechnique Fédérale de Lausanne (EPFL)
`karl.aberer@epfl.ch`

Abstract. Traditionally, data integration techniques involve central components, e.g., global schemas or ontologies, to overcome semantic heterogeneity for enabling transparent access to heterogeneous data sources. Today, however, with the explosion of machine processable formats in the Data Web, one cannot rely on global, centralized schemas anymore, as knowledge creation and consumption are getting more and more dynamic and decentralized.

Peer-to-peer data integration systems are a good example of this new breed of systems eliminating central semantic components and replacing them through decentralized processes of local schema alignment and query processing. As a result semantic interoperability becomes an emergent property of the system.

In this talk we will first survey recent developments in peer-to-peer data integration, illustrating which novel challenges and opportunities these systems introduce. We then illustrate of how semantic integration can be modeled as a self-organizing agreement process and present a probabilistic technique for distributed reasoning about semantic relationships. We show how to extend these methods from the problem of schema mapping to entity matching and how to consider trustworthiness of autonomous participants in this process. Finally we present some recent work on applying these principles in business applications.