# Peer-to-Peer Semantic Coordination

P. Bouquet[1], L. Serafini[2] and S. Zanobini[1]

[1] Department of Information and Communication Technology – University of Trento

Via Sommarive, 10 – 38050 Trento (Italy)

[2]ITC-IRST – Istituto per la Ricerca Scientifica e Tecnologica

Via Sommarive, 14 – 38050 Trento (Italy)

bouquet@dit.unitn.it serafini@itc.it zanobini@dit.unitn.it

Semantic coordination, namely the problem of finding an agreement on the meaning of heterogeneous schemas, is one of the key issues in the development of the Semantic Web (see [11, 10, 9, 8, 3, 5, 7, 1, 4, 6] for examples of proposed techniques).

In environments with more or less well-defined boundaries, like a corporate Intranet, the problem of semantic coordination can be addressed *a priori* by defining and using shared schemas (e.g. ontologies) throughout the entire organization[1]. However, in open environments, like the Semantic Web, this "centralized" approach to semantic coordination is not viable for several reasons, such as the difficulty of "negotiating" a shared model that suits the needs of all parties involved, the practical impossibility of maintaining such a model in a highly dynamic environment, the problem of finding a satisfactory mapping of pre-existing local schemas onto such a global model. In such a scenario, the problem of exchanging meaningful information across locally defined schemas (each possibly presupposing heterogeneous semantic models) seems particularly tough, as we cannot assume an *a priori* agreement, and therefore its solution requires a more dynamic and flexible form of coordination, which we call "peer-to-peer" semantic coordination.

In this paper, we address an important instance of the problem of peer-to-peer semantic coordination, namely the problem of coordinating hierarchical classifications (HCs). HCs are structures having the *explicit* purpose of organizing/classifying some kind of data (such as documents, records in a database, goods, activities, services). The main technical contribution is a logic–based algorithm, called CTXMATCH, for coordinating HCs. It takes in input two HCs $S$ and $S'$ and, for each pair of concepts $m \in S$ and $n \in S'$, returns their semantic relation. The relations we consider in this version of CTXMATCH are: $m$ is less general than $n$, $m$ is more general than $n$, $m$ is equivalent to $n$, $m$ is compatible with (possibly overlappings) $n$, and $m$ is incompatible with (i.e., disjoint from) $n$. The formal semantics of these relations will be made precise in the paper.

With respect to other approaches to semantic coordination proposed in the literature (often under different "headings", such as schema matching, ontology mapping, semantic integration), our approach is innovative in three main aspects: (1) we introduce a new method for making explicit the meaning of nodes in a HC (and in general, in structured semantic models) by combining three different types of knowledge, each of which has a specific role; (2) the result of applying this method is that we are able to produce a new representation of a HC, in which all relevant knowledge about the nodes (including their meaning in that specific HC) is encoded as a set of logical formulae; (3) mappings across nodes of two HCs are then deduced via logical reasoning, rather then derived through some more or less complex heuristic procedure, and thus can be assigned a clearly defined model-theoretic semantics. As we will show, this leads to a major conceptual shift, as the problem of semantic coordination between HCs is no longer tackled as a problem of computing linguistic or structural similarities (possibly with the help of a thesaurus and of other information about the type of arcs between nodes), but rather as a problem of deducing relations between formulae that represent the meaning of each concept in a given HC. This explains, for example, why our approach performs much better than other ones when two concepts are intuitively equivalent, but occur in

[1]But see [2] for a discussion of the drawbacks of this approach from the standpoint of Knowledge Management applications.

structurally very different HCs.

# References


[1] Sonia Bergamaschi, Silvana Castano, and Maurizio Vincini. Semantic integration of semistructured and structured data sources. *SIGMOD Record*, 28(1):54–59, 1999.

[2] M. Bonifacio, P. Bouquet, and P. Traverso. Enabling distributed knowledge management. managerial and technological implications. *Novatica and Informatik/Informatique*, III(1), 2002.

[3] Jeremy Carroll and Hewlett-Packard. Matching rdf graphs. In *Proc. in the first International Semantic Web Conference - ISWC 2002*, pages 5–15, 2002.

[4] A. Doan, J. Madhavan, P. Domingos, and A. Halevy. Learning to map between ontologies on the semantic web. In *Proceedings of WWW-2002, 11th International WWW Conference, Hawaii*, 2002.

[5] J. Euzenat and P. Valtchev. An integrativive proximity measure for ontology alignment. *Proceedings of the workshop on Semantic Integration*, October 2003.

[6] Ryutaro Ichisem, Hiedeaki Takeda, and Shinichi Honiden. Integrating multiple internet directories by instance–base learning. In *AI AND DATA INTEGRATION*, pages 22–28, 2003.

[7] Jayant Madhavan, Philip A. Bernstein, and Erhard Rahm. Generic schema matching with cupid. In *The VLDB Journal*, pages 49–58, 2001.

[8] Tova Milo and Sagit Zohar. Using schema matching to simplify heterogeneous data translation. In *Proc. 24th Int. Conf. Very Large Data Bases, VLDB*, pages 122–133, 24–27 1998.

[9] Marcello Pelillo, Kaleem Siddiqi, and Steven W. Zucker. Matching hierarchical structures using association graphs. *Lecture Notes in Computer Science*, 1407:3–??, 1998.

[10] Jason Tsong-Li Wang, Kaizhong Zhang, Karpjoo Jeong, and Dennis Shasha. A system for approximate tree matching. *Knowledge and Data Engineering*, 6(4):559–571, 1994.

[11] K. Zhang, J. T. L. Wang, and D. Shasha. On the editing distance between undirected acyclic graphs and related problems. In Z. Galil and E. Ukkonen, editors, *Proceedings of the 6th Annual Symposium on Combinatorial Pattern Matching*, volume 937, pages 395–407, Espoo, Finland, 1995. Springer-Verlag, Berlin.