Query Evaluation over Networks of Logic-based Mediators (Position Paper)

Carlo Meghini and Yannis Tzitzikas
CNR - ISTI, Pisa

1 The Framework

A network of information sources (NIS) consists of interconnected information sources, which can be of one of two types:

- *simple sources* are sources having a (possibly empty) theory and a database, interpreting the language of the theory;
- *mediators* are dataless sources, having a (possibly empty) theory and a number of mappings from the language of the theory to the language of some simple sources or mediators.

Information sources having both data and mappings can be considered as a pair formed by a mediator and a simple source, thus, without loss of generality, we can consider a NIS as consisting of just mediators and simple sources.
A network of information sources (NIS) can be characterized according to the following criteria:

- the numbers of mediators;
- the topology of the mappings, *i.e.* which mediators is mapped to which simple sources and other mediators;
- the types of the mappings. These can be described by means of triples $\langle C_g, C_l, \alpha \rangle$, where $\alpha \in \{sound, complete, exact\}$, and C_g and C_l can be as follows:
  - C_g is a mediator term and C_l is a query over a simple source, in which case the type of the mapping is global-as-view, or GAV;
  - C_l is a simple source term, and C_g is a query over the mediator, in which case the type of the mapping is local-as-view, or LAV.

2 Single Mediator

NISs with a single mediator are the most studied in the literature. They can be characterized as follows:

- there is only one mediator, obviously;
- the only existing mappings are between the mediator and each simple source;
- all mappings are of the same type, either GAV or LAV.

For the GAV approach, there exist algorithms and complexity results for a number of different data representation languages at the mediator (relational model with key and data dependency constraints, semantic data madels, the ER model). The LAV approach implies query answering based on views, and several cases of this problem have been studied and classified from the complexity point of view. It is very likely that, within the next few years, we will have a consistent body of results covering all possible instances of the single mediator setting.

3 Multiple mediators

NISs with several mediators have not yet been much studied, to the best of our knoweldge. Waht makes these systems appealing is their natural arising in the Internet, and also the fact that are very suitable for a peer-to-peer architecture, where various communities offer their data and knowledge so as to make up a global, decentralized information system, based on different languages and underlying theories. In this context, a mediator can be viewed as the gateway to the global information system for the community consisting of the simple sources served by the mediator. As such, a mediator may also be connected to other mediators, through a mapping of the respective languages. The mappings amongst a mediator and its local sources (mediator-source mappings) support integration within the same linguistic and ontological community; the mapping amongst mediators (mediator-mediator mapping) support integration with communities with different languages or ontologies.
The mediator-source kind of mapping, in a peer-to-peer architecture, is most naturally seen as a LAV mapping: the mediator acts as the keeper of the language and the ontology of the community, and each source belonging to that community describes its data in terms of its mediator's language. In this way, the information system is able to scale in an effective manner, as addition and removal of a simple sources can be done without any impact on the other sources of the network.
The mediator-mediator kind of mapping is most naturally seen as a GAV mapping. Here, scalability is achieved by giving a mediator the possibility of connecting its terms with those of any other mediator of the network. This implies the possibility of loops in the graph representing mediators' connections.

4 Query Evaluation

The problem that we intend to address within the above described framework is that of query evaluation. The idea is that a query can be posed at any node of the network, in any language. The query is then propagated until it reaches a mediator that "understands" it; this mediator becomes the coordinator of the query evaluation.
Roughly, the task of the mediator in evaluating the query can be split into two independent subtasks:

- the *data collection* task, in which the mediator collects the qualifying data from the simple sources mapped to it; and
- the *query routing* task, in which the mediator rewrites the query for each of the mediators for which it has mappings, and forwards the re-written queries to the appropriate mediators. Query evaluation then proceeds recursively, under the control of the involved mediators, until some termination condition is reached.

In this scenario, Description Logics (DLs) are appropriate for representing mediators' ontologies. As it is well-known, the vast majority of conceptual data modelling formalisms can be captured

by DLs, and a wide range of results exist for the main reasoning tasks. Mappings, on the other hand, are most appropriatedly expressed as logic programs. The resulting system would thus look like an hybrid, featuring DL-based ontologies mapped through logic programs to other DL-based ontologies. This kind of hybrid systems have been investigated in the last decade. Datalog (or less expressive) queries are posed against this system.

We have started investigating the evaluation of queries in NISs with more than one mediator, addressing the propositional case, *i.e.* the case in which ontologies are simple term taxonomies, and the mappings are of the GAV, sound type, expressed as propositional Horn clauses. We have various algorithms for query evaluation, depending on a few underlying assumptions. We plan to move towards more expressive formalisms, starting considering DLs as ontology representation languages.