Conceptual Concistency Management in Multidatabases: The FEvoS Framework

Nick Pittas, Andrew.C.Jones and W.A.Gray
Department of Computer Science, Cardiff University,
PO Box 916, Cardiff CF24 3XF, UK.
N.Pittas@cs.cf.ac.uk

Abstract

While a number of different approaches exist to facilitate the interoperation of distributed, heterogeneous and autonomous databases, a systematic framework to assist with the evolution process of federated systems has yet to be presented. A primary hurdle in the life cycle of an interoperable database system is the effort, and therefore the cost, that is associated with its maintenance following its initial launch. In this paper we present a framework that can be used for the algorithmic detection of inconsistencies in evolving multidatabase systems, at both structural and semantic levels. Such framework forms the basis of a prototype system. We focus on the detection and reporting of incongruities that occur in the way that local entities are ontologically committed to global ones. Our proposal is founded on the exploitation of intensional sources' data in the creation of a construct that represents explicitly the ontological commitment of local entities onto federal concepts, which we refer to as *Conceptual Links*.

Keywords: Interoperable Systems, Metadata Exploitation, Conceptual Consistency Management, Federated Systems Evolution

Recent developments, driven by the dramatic proliferation of databases and the advances in wide-area networking, such as the ones envisaged in [10], [2] or reported by [1], [3] presume the existence of federations of non-trivial size. Indeed, current research reports a number of different approaches [4], [6], [8], that facilitate the **interoperation** of distributed, heterogeneous and autonomous databases but a systematic framework to assist with their evolution process has yet to be presented. One of the fundamental difficulties associated with large-scale database interoperation is the labour intensiveness, and therefore the cost, required for detecting and reconciling discrepancies that arise as the result of the evolution process of federated systems in order to maintain the federation in a consistent state after its initial launch [5]. Such difficulties are already major issues, although not so profound due to lack of scale, in small networks of interoperating and heterogeneous data sources [1]. The main focus of *FEvoS*[1] [7] is in the provision of an algorithmic

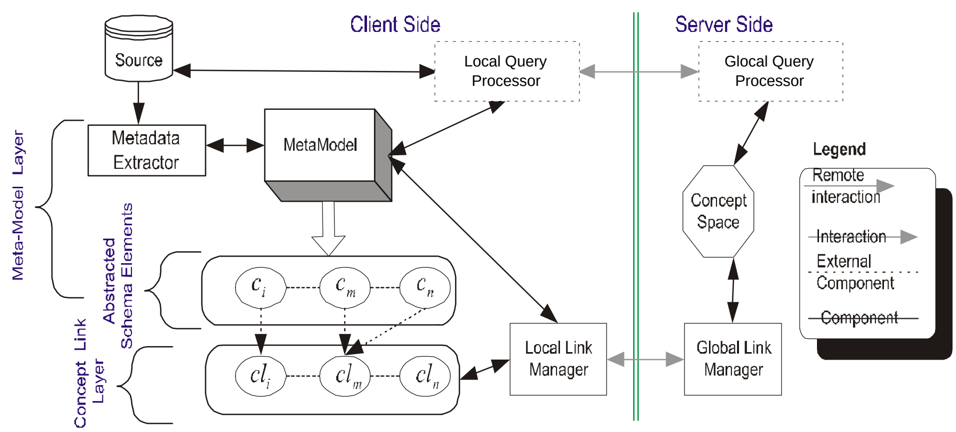

Figure 1: Main Prototype System Architecture.

framework in support of the conceptual evolution process of multidatabase systems. More specifically *FEvoS* exploits intensional component sources' data and semantic information, as expressed in domain ontologies [9], in order to dynamically detect and report incongruities that arise in the way that local entities are ontologically committed onto federal ones. Incongruities discovered are reported back to affected sources, along with a report stating their nature and potential causes, for reconciliation.

The main architecture of our approach is depicted in figure 1. *FEvoS* employs a 4-layered knowledge organisation. The first layer is composed of the

[1]FEvoS is a paraphrase of Phoebus, the ancient Greek and Roman deity for music and poetry, and stands for **F**ederation **Evo**lution **S**ystem.

component sources which remain unaltered by the introduction of the evolution framework. The second layer is the metamodel, of which the purpose is two-fold. Firstly it provides an abstraction for intensional source schema elements properties and semantics to be captured in a way independent of the data organisation of its originating source; secondly it acts as the means of persistently capturing the source's current schema state and thus enabling the retrieval of previous and current schema element structural states, therefore making feasible the comparison between them and the identification of changes. The third layer (the *conceptual links* layer) provides the means for making explicit how local entities are ontologically committed onto global ones. It does so by enlisting all schema attributes necessary to represent a global concept(s). Its formulation is solely based on the abstraction of sources' schema elements as pertained in the metamodel. The fourth layer, the *Concept Space* is the framework's ontology and acts as an explicit specification of global concepts and their semantic domains. Concepts in the *Concept Space* are represented by a graph with two type of edges, association and aggregation links. The Concept Space presents a highly structured construct that enables the detection and reporting of discrepancies arising from changes in the representation or semantics of federal system entities. Furthermore it allows for the mapping of such incongruities onto affected *Conceptual Links*.

Alterations which render the ontological commitment inconsistent may occur at two distinct functional levels in multidatabase systems:

–**Global**: Changes at a global level reflect evolving global user or application requirements. Such changes assume the form of discrepancies in the way that a multidatabase system perceives global concepts.

–**Local**: Local changes are propagated as a direct result of schematic changes in component sources or changes in the implicit semantics represented by a source's intensional elements. They are the product of evolving local user requirements or system-level enhancements (e.g. schema re-organisation).

Dynamic change detection in our framework takes place in the form of two algorithms, comprising a distributed one, the $FedEvo_{Global}$ [7] algorithm and the $FedEvo_{Local}$ [7] algorithm. The $FedEvo_{Global}$ algorithm works by comparing the current state of the *Concept Space* with the one persistently stored at the time of the last invocation. It considers how changes at a global level affect ontological commitment. Its end product is the set of affected *Conceptual Links* and a report stating the nature of the change which is selectively forwarded to only affected sources. In a similar manner the $FedEvo_{Local}$ algorithm founds its operation in the comparison of the immediately previous and current metamodel states. The algorithm is able to distinguish among introduced, deleted and evolved source schema elements by employing a fuzzy logic approach based on schematic elements' intensional properties. Its product is again the set of conceptual links that are affected by the schematic change(s) and provides complete coverage of schematic conflicts that can occur in multidatabases as enlisted in a widely accepted classification in [4].

The potential benefit that such framework has to offer is the alleviation of the need for human based discrepancies discovery, an otherwise tedious process, and therefore a significant reduction in the cost and labour-intensiveness of federated systems evolution. We are currently in the process of quantifying such benefits in a system of appropriate scale, such as we have described in [3]. While currently the framework deals only with how changes affect the federation at a conceptual level it could be extended to support discovery among local and global constraint mappings and thus enhance its completeness. Finally another interesting domain that we believe such framework can be successfully applied to is the enhancement of source *"wrappers"* with self diagnostic capabilities.

References

[1] Athman Bouguettaya et al. Implementation of interoperability in large multidatabases. *In Proceedings of RIDE-IMS93*, pages 55–60, 1993.

[2] Sharma Chakravarthy. Just in time information: to push or not to push. *In Proceedings of the 17th British National Conference on Databases (BNCOD17), Exeter,UK*, pages 19–20, July 2000.

[3] Andrew.C. Jones, X. Xu, Nick Pittas, W.A. Gray, N.J. Fiddian, R.J. White, J.S. Robinson, F.A. Bisby, and S.M. Brandt. Spice: A flexible architecture for integrating autonomous databases to comprise a distributed catalogue of life. *Database and Expert Systems Applications (DEXA2000), Greenwich,UK*, pages 981–992, 2000.

[4] W. Kim and J. Seo. Classifying schematic and data heterogeneity in multidatabase systems. *IEEE Computer*, pages 12–18, December 1991.

[5] B.P. Lientz. Issues in software maintenance. *Computing Surveys*, 15:271–278, 1983.

[6] W. Litwin. From database systems to multidatabase systems: Why and how. In *Proceedings of the 6th British National Conference on Databases(BNCOD6)*, pages 161–188, Cardiff, U.K., July 1988.

[7] Nick Pittas, A.C.Jones, and W.A.Gray. Evolution support in large-scale interoperable systems: a metadata driven approach. In *Proceedings of the 12th Australasian Database Conference*, pages 161–168. IEEE CS Press, 2001.

[8] A. P. Sheth and J. P. Larson. Federated database systems for managing distributed, heterogeneous, and autonomous databases. *ACM Computing Surveys*, 22(3):183–236, September 1990.

[9] G. Wiederhold. Interoperation, mediation and ontologies. In *International Symposium on Fifth Generation Computer Systems(FGCS94), Workshop on Heterogeneous Cooperative Knowledge-Bases*, pages 33–48, Tokyo, Japan, December 1994.

[10] Gio Wiederhold. Precision in procesing data from heterogeneous resources. *In Proceedings of the 17th British National Conference on Databases (BNCOD17), Exeter,UK*, pages 1–18, July 2000.