

Report on the EFIS 2001 Workshop

Background

Engineering of Federated Information Systems (FIS) is a research area, which addresses the systematic development of interoperation solutions for autonomous, distributed heterogeneous systems covering both database and non-database information sources. As many examples of information systems for business, science and personal use show, there is an increasing need to make information available via some kind of network (enterprise networks or even the WWW), not only for “global” applications, but also for local applications using semantically related information.

This workshop is the fourth of a series of international workshops on Engineering Federated Information Systems. This fourth workshop (EFIS 2001) strengthened the exchange of ideas and experiences between academia and industrial practice. With this goal in mind, EFIS 2001 was included as part of the Information Federation Week (IFW) in Berlin, which combined the EFIS workshop with an industrial day (INTEGRA) and the national German workshop on Federated Database Systems. Reports on the previous EFIS workshops are available in ACM SIGMOD Record 26(4), December 1997, ACM SIGMOD Record 28(3), September 1999 and ACM SIGMOD Record 29 (4), December 2000. The workshop homepage is: http://cic.cs.tu-berlin.de/~ifw2001/efis/engl/index_efis.html. Information on the proceedings of the workshop may be found at: <http://www.iospress.nl/site/html/pis.html>.

Invited Presentations

There were two invited presentations, one from academia and one from industry:

- A Voisard (FU Berlin): A Framework for Integrating the Documentation of Cooperative Work

- L.P Kurdelski (BEA): Enterprise Application Integration (EAI) and Inter-Enterprise Integration (IEI) – The third integration-generation based on application servers

The growth of collaborative working on projects at both national and international levels has given rise to the issue of integration of documentation. The first invited presentation described work that had been carried out to address this problem for a particular project in the area of spatial information systems in geology. The presentation set out a framework for allowing documentation of cooperative work to be meaningfully shared. It showed how documentation could be differentiated as “what”, “how” and “where” and integrated into a single database.

The second invited presentation was an industrial presentation, which addressed middleware deployment and use of tools/frameworks. Commercial middleware (e.g. in EAI architecture) is now available and it ensures a high level of technical independence for FIS. The subsequent INTEGRA workshop of the IFW gave further evidence of this. Semantic aspects however have not yet been addressed by commercial vendors.

Technical Papers

The following papers were selected for presentation:

- [1] Shallehn et al (Univ. Magdeburg, Germany): Extensible Grouping and Aggregation for Data Reconciliation
- [2] Pittas et al (Cardiff Univ., UK): Metadata Exploitation in Support of Federated Database Systems Evolution
- [3] Thiran & Hainaut (Univ Namur, Belgium): Evolving Hybrid Distributed Databases : Architecture and Methodology
- [4] Kamur and Roantree (Dublin City Univ., Eire): Using Stored Behaviour in Object-Oriented Databases

- [5] Rodriguez et al (Univ Catalonia, Spain): On Operations along the Generalization/Specialization Dimension

- [6] Wyss et al (Indiana Univ. USA): Augmenting SQL with Dynamic Typing to support Interoperability in a Relational Federation

- [7] James and Salem (Coventry Univ., UK): Classification and Resolution of Behavioural Conflicts in Federated Information Systems

- [8] Altareva and Conrad (Munich Univ, Germany): The Problem of Uncertainty and Database Integration

Main Topics

The workshop had an emphasis on evolution of metadata. This topic had been identified in previous workshops as an important topic for future work.

Conforming operations, which change schema structure to a predefined pattern, were discussed. [5] presents a set of conforming operations for generalisation and specialisation patterns. Such operations are seen as providing a useful contribution in multi-schema systems. The issue of maintenance of interoperable systems was considered. [2] introduces a formal framework and algorithms which detect and classify the effects of change in evolving multi-database environments. Changes at both the local and the global level are handled.

Another topic emphasised at the workshop was the topic of integrating behaviour [4, 7]. This too had previously been identified as an area for future work. Object-oriented techniques are the preferred choice for addressing the problem of behaviour integration. However, the existing standard for object-oriented databases (ODMG) is considered too restrictive and should be extended with further behavioural support [4]. A behaviour taxonomy for object-oriented databases is provided in [4] together with an outline of an extended ODMG meta model. [7] gives the first steps for a

classification of behavioural conflicts and proposes some strategies for their resolution. Whilst [7] and [4] concentrate mainly on the meta level of behaviour, it is recognised that previous work on behavioural integration at the level of behaviour description using statecharts (Frank and Eder 1999) or petri-net like notations (Preuner et al 2001) will form a useful contribution to a holistic solution to behavioural integration.

A more traditional topic was that of developing extensions to multi-database query languages [1,6]. [1] proposes an extension to an SQL-like query language by introducing concepts for grouping and aggregation. These extensions address advanced requirements regarding data reconciliation in integration processes. [6] proposes to add to SQL the ability to dynamically restructure relational data among databases, based on the data and meta data of the input relations. Thus data may become metadata and metadata may become data.

Integration methodology was another issue discussed [3,8]. The problem of meeting new requirements in a legacy system context was considered. [3] proposes that both forward and reverse engineering are equally important in developing a global schema. An architecture is proposed which combines forward and backward processes and which is based on wrappers and mediators. This topic has been discussed in previous EFIS workshops [van den Heuvel et al 2000]. In this workshop, it was emphasised that reverse and forward processes are tightly bound.

Re-engineering is useful in building a FIS when strong knowledge of the underlying systems can be exploited. In such cases it can be introduced as a deterministic step. However the schemas of component systems may be incomplete or the semantics of legacy systems may not be completely known and documentation may be provided in diverse form. In these situations data mining techniques may be useful to discover the necessary information for reverse engineering. Such techniques, which may analyse program code, schemas and data, deliver data with uncertainty or limited confidence. [8]

suggests some first steps for a method of integration that takes into account uncertainty of inter-schema data relationships.

Requirements of specific application domains were not discussed in the workshop

Discussion

In the case of most presentations at the workshop, database systems (or at least databases with well structured schemas) were considered as sources of FIS. In practice this seems to be only a special case because often databases with incomplete or partly unknown schemas (e.g. coming from legacy systems) have to be integrated. Furthermore, semi-structured data from other sources (e.g. the web) need to be integrated into an FIS as well. Some problems arising from partially unknown schemas or from semi-structured data are discussed in [8].

The observation that most contributions still addressed database-oriented issues and also the mainly industry-driven trend of middleware solutions to the integration problem, led to the discussion of semantics in the context of FIS management and evolution. The middleware community see XML as an 'all-healing' medicine within an integration framework which includes tools for communication and interoperation. However the task of semantic integration remains in the applications which are built on the framework. Integrated solutions for information integration and middleware platforms were not presented at the workshop. The question of metadata construction for an evolutionary development of FIS was addressed in the discussion but not in the talks and papers.

Identified Topics for future work

The observations made at the workshop and our understanding of the whole area of FIS yields a number of open questions to be discussed in the future, possibly at the next EFIS workshop in 2003. These are:

- The role of metadata as part of information federations
- Relating semantics to middleware platforms

- Evolution of FIS, based on strongly autonomous environments
- Dealing with uncertainty, particularly identifying appropriate thresholds for deciding on similarity of heterogeneous information models as well as resolving conflicts.
- Improved support for integrating behaviour
- Federation architecture is based on a mediator hierarchy. Defining the nodes in the mediator hierarchy is an important issue that has yet to be worked out
- Tool support for building FIS and interoperability of such tools

Outlook

The fifth workshop of the EFIS series will be held at Coventry University, UK in the summer of 2003. It may be associated or co-located with the British National Conference on Databases, which will also be held at Coventry University in 2003.

References

- Frank H, Eder J, Towards an Automatic Integration of Statecharts, Proc. 18th International Conference on Conceptual Modeling, Paris, France, LNCS 1728, Springer-Verlag, pp.430-444, November 1999
- Preuner G, Conrad S, and Schrefl M, View Integration of Behaviour in Object-oriented Databases, Data and Knowledge Engineering, 36(1), pp 153-183, 2001
- Van de Heuvel, Hasselbring W and Papazoglou M, Top-down Enterprise Application Integration with Reference Models, Australian Journal of Information Systems, 8, 1, pp126-136, September 2000

¹ Univ. Munich, Germany, Conrad@dbs.informatik.uni-muenchen.de

² Univ. Oldenburg, Germany, hasselbring@informatik.uni-oldenburg.de

³ Coventry University, UK, a.james@coventry.ac.uk

⁴ DCU, Ireland, dalenk@compapp.dcu.ie

⁵ TU Berlin, Germany, rkutsche@cs.tu-berlin.de

⁶ Univ. Namur Belgium, pth@info.fundp.ac.be