VALIDATION, VERIFICATION AND INTEGRITY ISSUES IN EXPERT AND DATABASE SYSTEMS

T. Bench-Capon[1], *D. Castelli*[2], *F. Coenen*[1], *L. Devendeville-Brisoux*[3], *B. Eaglestone*[4], *N. Fiddian*[5], *A. Gray*[5], *A. Ligeza*[6] *and A. Vermesan*[7]

1) Department of Computer Science, University of Liverpool, UK

2) Istituto di Elaborazione dell'Informazione, Consiglio Nazionale delle Ricerche, Pisa, Italy

3) Centre de Recherche en Informatique de Lens, Universit d'Artois, France

4) Department of Information Studies, University of Sheffield, UK

5) Department of Computer Science, Cardiff University, UK

6) Institute of Automatics, Krakow, Poland

7) Department for Strategic Research, Det Norske Veritas, Olso, Norway

Abstract

In August 1998 an international workshop on validation, verification and integrity issues of expert and database systems was held in Vienna in conjunction with the Ninth Conference on Database and Expert Systems Applications (DEXA'98). This paper reports on the results of this workshop and summarises the issues identified for future research in this area. Details of the presented papers and some additional information can be found at the workshop WWW site `http://www.csc.liv.ac.uk/~frans/dexaWorkshop.html`.

1 INTRODUCTION

Validation, verification and integrity (VV&I) respectively concern the truthfulness of a system with respect to its problem domain, the correctness of its construction, and maintaining the validity of the evolving system throughout its life. This distinction, akin to the differences in traditional software engineering, between these three terms is neatly captured in Bohem's definitions of validation and verification [Bohem 81] and the additional definition of integrity by Eaglestone and Ridley [Eaglestone & Ridley 98]):

Validation: Are we building the right product?

Verification: Are we building the product right?

integrity: Are we keeping (maintaining) the product right?

In practice, integrity mechanisms fall short of this, since they ensure only self-consistency and therefore maintain plausibility rather than truthfulness.

Verification and validation have long been a focus for research within the expert systems community [Coenen 98]. Although the technology of Rule-Based Systems has become more and more "common place", the technology is still not well-accepted by industrial engineers. Further, its "correct" usage requires much intuition and domain experience, and knowledge acquisition still constitutes a bottleneck for many potential applications. A serious problem consists in the fact that verification of theoretical properties remains problematic. Hence, research toward more formal design and verification of rule based systems constitutes an important issue for assuring their correctness, reliability and safety. Research work on Expert Systems V&V is also stimulated [Vermesan 98], by the practical application of theoretical results to provide the tools and techniques for certification of knowledge-based software.

In contrast, the terms "validation" and "verification" are rarely used in the database literature [Eaglestone and Ridley 98]. Instead, researchers have focused on the dynamic nature of database systems and the problem of maintaining database integrity. (However, it is noted that V&V of computer systems in the broader sense, involving testing of their database component, is an active area in the software engineering community.) In addition and contrary to expert systems, data base systems (and especially relational DB systems) are a matured data manipulation technology employing a widely accepted intuitive knowledge representation in tabular form. It seems advantageous to make use of elements of this technology for simplifying certain operations concerning rule based expert systems. Note that from a practical point of view any row of a relational DB table can be considered as a rule, provided that at least one attribute has been selected as an output (and there is a so-called 'functional dependency' allowing for determination of the value of this attribute on the base of some other attributes). Thus, it seems that merging elements of the two technologies can constitute an interesting research area of potential practical importance.

"Integrity" is of course a database issue, however, if the meaning of the term is widened, as suggested by Eaglestone and Ridley, to include maintenance then this becomes an issue for both KBS and databases systems. The maintenance of KBS, an issue that has long been researched by Bench-Cpaon and Coenen [Coenen and Bench-Capon 1993], in many cases encompases a subset of the operations required for more general V&V of expert systems. When maintaining/updating the "fact base" associated with a KB, there is a much closer correspondence with the term "integrity" as used in the database context.

The aim of the workshop was therefore to review the state of the art with respect to validation, verification and integrity (VV&I) issues in both communities with the objective of identifying overlap and areas where transfer of results between communities can be mutually beneficial. The proceedings of the workshop are available in [Wagner 98]. The workshop reflected both the database and expert systems views of VV&I, and covered the following topics:

The expert systems view (overviewed in [Coenen 98]):

- Industrial use of validation and verification;
- Verification and validation of rule-based systems;
- Verification of multi agent knowledge systems;
- Ontology-based support for validation and verification.

The database systems view (overviewed in [Eaglestone and Ridley 98]):

- Integrity constraints in distributed database systems;
- Integrity constraints on automatically classified data objects;
- Verification of database schema transformations.

2 THE EXPERT-SYSTEMS VIEW

Expert-systems related work presented addressed the practical application of V&V results and theoretical work using different knowledge-based system paradigms.

Vermesan reviewed work on the practical use of V&V results for certification of expert systems [Vermesan 98]. She identified this application as probably the most advanced use of V&V methods, techniques and knowledge so far developed. Knowledge base (KB) certification uses a number of different V&V methods such as static analysis, testing, inspection, and modeling to assess the quality of the KB component. Although these methods are currently used in expert system development, it is not yet clear how they would be applied by an independent certification organization to arrive at a pass/fail decision. Her paper identifies application areas directly connected with industrial needs and discusses the V&V as component of KB certification, advocating a combination of static and dynamic testing; the former to remove anomalies and errors prior to dynamic testing.

Ligeza's presentation concerned V&V of tabular rule-based systems [Ligeza 98]. Ligeza noted that a tabular rule-based system has a uniform structure not unlike a relational database table; thus it was argued that the paper investigated rule base V&V issues using a relation DB algebraic knowledge representation. The main difference from pure relational DB systems consists in perceiving the particular records as rules rather than static data records. This is done by distinguishing in each rule-record the precondition part and the decision part; further, not all the values of attributes have to be specified by a single value. Ligeza also presented a taxonomy of issues concerning formal verification, and demonstrated that several theoretical properties can be analysed through simple algebraic operations instead of logical inference. Ligeza then went on to suggest that algebraic notation, on the lines of the relational DB model, was both simpler and more intuitive than pure logic, and therefore more acceptable to practitioners.

V&V of multiple agent knowledge-based systems was addressed in [O'Leary 98]. This work described extensions to verification tests to systems with multiple autonomous agent knowledge bases. The techniques were primarily to detect unique inter-agent anomalies generated as a system moves from a single agent to multiple agents.

Original work on stochastic search for the validation of first-order knowledge-based systems was reported in [Brisoux, Grgoire and Sais 98]. It allows for both logical consistency and inconsistency constraints to be checked in large first-order knowledge bases. It is based on powerful heuristics about the trace of local search methods when they fail to prove the consistency of large propositional knowledge bases. A partial instantiation schema is proposed that avoids the combinatorial explosion that could occur when a full instantiation of a first-order knowledge base into a flat corresponding propositional counterpart is undertaken. Accordingly, a depth-limited consistency and inconsistency checking procedure is proposed and successfully experimented.

Bench-Capon [Bench-Capon 98] discussed the ways in which ontologies which conceptualise the problem domain can be used to support V&V of knowledge-based systems. Bench-Capon observed that the added coherence of an explicit domain conceptualisation gives a more motivated and foccussed role for experts who are "signing off" a KB, reduces the scope for subjective interpretation which can mask errors, and provides the possibility of a test harness.

3 THE DATABASE VIEW

The database-related work concerned integrity issues and verification of schema transformations/database design.

Ibrahim et al [Ibrahim, Gray and Fiddian 98] described a technique for determining an optimal set of constraints for validating a distributed database created by fragmenting a centralised database. Their presentation covered integrity checking in a centralised database in order to maintain the consistency of the database. They then showed that if such a database is distributed by fragmentation it is normally no longer efficient to use the same integrity constraints. They then showed how knowledge about the fragmentation used on the data can be used to create a new set of integrity constraints that can be distributed to the sites of distributed system and improve the integrity checking in the distributed system as they reflect the data held at the site.

Millns and Eaglestone's work concerned the use of neural networks embedded in database systems as automatic classifiers. They identified the necessity to bound the percentage of misclassified objects in certain critical applications and described an integrity constraint to enforce this bounding. The constraint was based on deducing a relationship between probabilities returned by probabilistic neural networks and the likely percentage of misclassification.

Finally, Castelli and Pisani presented a formal approach based on schema transformations for ensuring the correctness of database schema design. Each transformation has associated conditions which, if verified, guarantee the correctness of the design. In constrast to similar approaches, in this proposal the set of possible transformations is not fixed, but can be enlarged dynamically to best fit the application design needs.

4 RESULTS

In seeking the overlap and possible transference of results, the workshop identified the following (fuzzy) correspondences between database and expert systems architectures (Figure 1).

The following observations were made regarding this overlap:

- The ontology component is "an explicit conceptualisation of the domain" [Bench-Capon'98]. However, though an ontology is at present not usually defined for either type of system, some delegates identified considerable advantages in doing so for both database and expert systems. Specifically, an ontology provides an agreed representation of the Universe of Discourse and hence support for validation and verification based upon real world semantics.

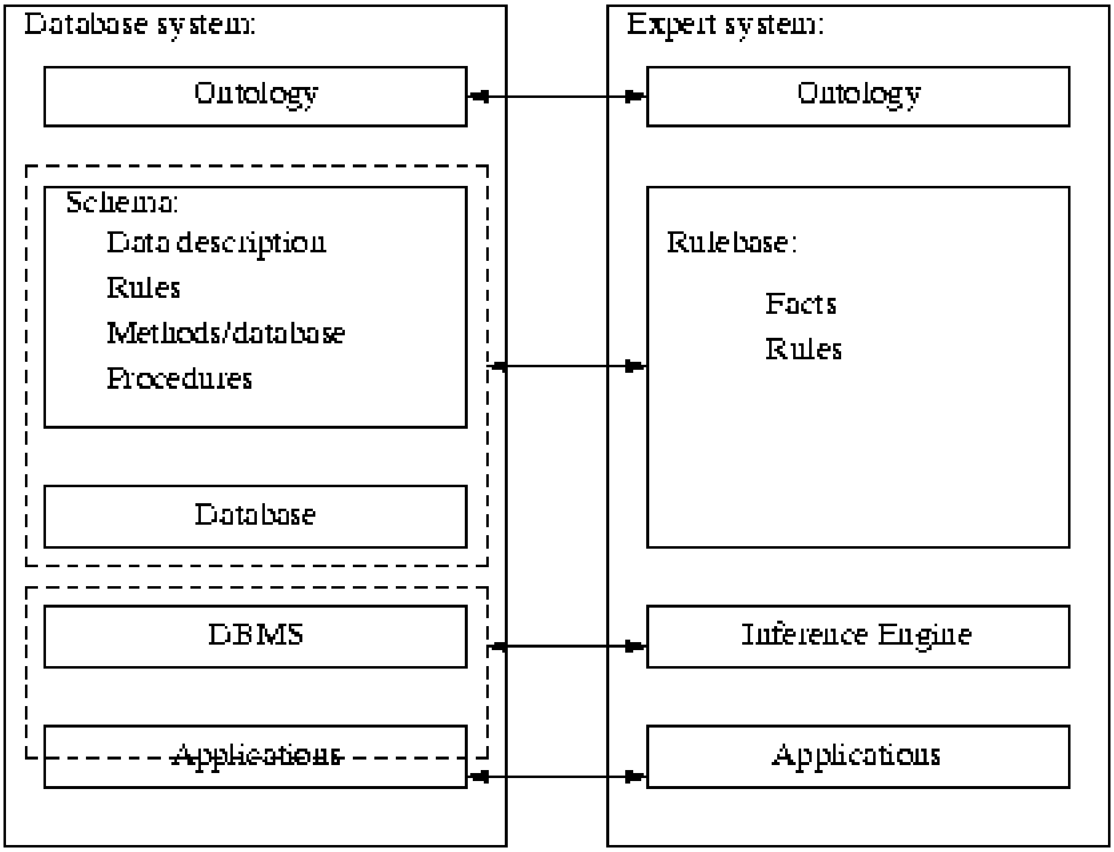

Figure 1: Correspondences between expert and database systems

- Correspondence between schema/database and rule base is most apparent for relational databases, since they share a common theoretical foundation with expert systems, i.e. clausal form. This has been exploited, for example in deductive databases where relational DBMSs provide back-end fact storage. However, this correspondence has not been exploited practically to any extent for V&V of database systems or for integrity maintenance of evolving expert systems.
- The emerging database technology, i.e. object and object-relational, has resulted in a divergence, since the new technology lacks the rigorous formal foundations shared by relational and expert systems.
- Validation and verification of the DBMS and Inference engine is distinct from but a necessary prerequisite to validation and verification of the (design of a) database or expert system. The current assumption made by system developers that this has taken place may well be ill founded in many cases.
- The overlap between the two technologies is blurred by the existence of hybrid technologies which borrow from both. These include deductive databases, constraint databases and active databases, all of which are supported by significant rule sets of a similar nature to that found in expert system rule bases. In theory, such hybrid products should be attractive because they derive benefits from both database and expert systems technologies, but in practice they remain largely “academic toys”.
- It may also be a point that Case-Based Reasoning systems may constitute and intermediate link integrating RDBS and RBS methodologies into perhaps more sophisticated models of knowledge.

5 CONCLUSIONS

The 1st International Workshop on Validation, Verification and Integrity Issues as a whole was considered to be a great success in that it succeeded in bringing representatives from the two communities together and facilitated the interchange of ideas between the two groups. Clear areas of overlap were identified between expert and database systems technologies and also a shared practical need for VV&I tools and techniques.

It was also observed [Ligeza] that relational DBs have a well established model and matured design methodology, while in the case of knowledge-based systems no such well established model exists, e.g. a widely accepted knowledge representation standard and inference engine on the lines of (say) SQL. Further, it was noted that this was especially the case with respect to rule base design and verification procedures. Consequently the database systems community might have a lot to offer the developers of Expert Systems.

Despite the above it was apparent that there is currently little common research effort or shared results. It was thus the broad view of the workshop delegates that dialogue and collaborations between the two communities was long overdue and that the DEXA'98 Workshop was the start of this "long overdue" liaison. To further facilitate the desired dialogue it was acknowledged that (1) the remit of the well established V&V "special interest group" within the Expert Systems community should be broadened to include database systems issues and (2) that the database practitioners would attempt to identify and integrate relevant results in the database systems area. With respect to (1) this is already being addressed in that the CFP for EUROVAV'99 (The principal international expert system V&V biannual symposium) includes database integrity, and database and expert systems integration V&V issues, in its "papers are invited (but not restricted to) the following topics:" list. Consideration is also currently being directed to (2). What ever the case it is intended to hold the 2nd International Workshop on Validation, Verification and Integrity Issues at DEXA'2000.

References

[1] Proceedings DEXA 98 Ninth International Workshop on Database and Expert Systems Applications, August 260-28, 1988, Vienna, Austria, Ed. R.R. Wagner, IEEE Computer Society, 1998.

[2] Bench-Capon, T.J.M., The Role of Ontologies in the Verification and Validation of Knowledge Based Systems, in [1].

[3] Brisoux, L., Gregoire, E. and Sais, L., Validation of Knowledge-Based Systems by Means of Stochastic Search, in [1].

[4] Castelli, D. and Pisani, S., Ensuring Correctness of Personalized Schema Refinement Transformations, in [1].

[5] Coenen, F., Verification and Validation Issues in Expert and Database Systems: The Expert Systems Perspective, in [1].

[6] Coenen, F. and Bench-Capon, T. (1993). Maintenance of Knowledge Based Systems: Theory, Tools and Techniques. Academic Press, London.

[7] Eaglestone, B. and Ridley, M., Verification, Validation and Integrity Issues in Expert and Database Systems: The Database Perspective, in [1].

[8] Ibrahim, H., Gray, W. A., and Fiddian, N.J., Optimizing Fragment Constraints, in [1].

[9] Ligeza, A., Towards Logical Analysis of Tabular Rule-Based Systems, in [1].

[10] Millns, I. and Eaglestone, B., An Integrity Constraint for Database Systems Containing Embedded Neural Networks, in [1].

[11] Vermesan, A.I., Software Certification for Industry - Verification and Validation Issues in Expert Systems, in [1].