AI Magazine Volume 10 Number 4 (1989) (© AAAI)

Databases in Large AI Systems

O. Friesen and F. Golshani

■ *Databases are at the heart of most real-world knowledge base systems. The management and effective use of these databases will be the limiting factors in our ability to build ever more complex AI systems. This article reports on a workshop that explored how databases and their associated technologies can best be used in the development of large AI applications.*

On 26 August 1988, approximately 50 people assembled at a workshop sponsored by the American Association for Artificial Intelligence (AAAI) to share ideas about the use of database management techniques in large AI systems. The organizers of the workshop were Forouzan Golshani, Department of Computer Science, Arizona State University, chairman; Ron Ashany, Department of Computer Science, University of California, Berkeley; Michael Brodie, GTE Labs., Waltham, Massachusetts; Oris Friesen, Bull HN Information Systems, Phoenix, Arizona; Sara Graves, Department of Computer Science, University of Alabama, Huntsville; and Carlo Zaniolo, MCC, Austin, Texas.

Forty-two papers were received for consideration. Of these papers, nine were presented at the workshop. In addition to the presented papers, a panel was convened to discuss the viability of database management in AI systems. Ten additional papers were included in the workshop proceedings.

Invitations were sent to 50 people to attend the workshop; 40 persons accepted and participated. An additional 10 people were allowed to attend the workshop as observers.

This article presents (1) a summary of each presentation and the panelists' statements, (2) a selective list of some of the more salient points raised during the discussions by various workshop participants, and (3) some concluding remarks.

Summary

The first session was chaired by Forouzan Golshani. In *A KBMS Architecture for Many-to-Many Coupling of Knowledge Systems to Databases,* James Davis of NCR described an architecture that attempts to solve the problem of conventional database management system (DBMS) support for multiple knowledge-based systems. It incorporates an enhanced knowledge dictionary (EDICT) that is stored as an extension of the DBMS data dictionary. He described an approach to reconciling the knowledge base levels, defined by Mark Fox, with the semantic views often used in database systems (for example, enterprise view and external view), which generated a good deal of discussion.

In *Intelligent Databases: Approach and Directions,* Anil Nigam of IBM Yorktown predicted that future developments in the area of intelligent databases would primarily occur in the spheres of new functions (for example, semantic proximity, qualitative responses, and multiple related databases) and new interfaces (for example, exploratory interrogation, interesting objects, and feedback of the query's supposed meaning), with system improvements playing a relatively minor role. He went on to describe a representation scheme—based on KL-ONE—called KL-DB; KL-DB extends the relational model to include semantics.

In *Toward a Synergism of Expert Systems and Databases*, Shel Finkelstein of IBM Almaden Research Center addressed the five research areas of integration, representation, language, search mechanisms, and binding, especially as they relate to the Starburst Project. Although Starburst uses structured query language (SQL) and the relational model, it relies on a custom-made DBMS (a feature that is quite common in the AI community).

The next session was chaired by Michael Brodie. In the first paper, *Security in Large AI Systems,* Bhavani Thuraisingham of Honeywell discussed the applicability of discretionary access controls (used widely in the database world) for AI systems. She categorized three types of discretionary rules deemed appropriate for AI systems: application-specific rules, application-independent rules, and inference rules. She also described a hypothetical architecture that would support such a discretionary policy.

In *The ISR: A Database for Symbolic Processing in Computer Vision,* John Brolio of the University of Massachusetts described an intermediate symbolic representation (ISR), which is based on the traditional database management methodologies. ISR is used to mediate access to massive quantities of vision data and provide an interface to the higher-level inference processes responsible for deriving the meaning of an image.

In *CYC: Consistency, Distribution, and Cooperation,* Chris Maeda of MCC addressed the problems and costs of providing consistency and integrity in very large knowledge bases. He posed a number of interesting considerations that have only been dealt with in a cursory manner by the AI community. For example, data-integrity problems are often resolved post facto by unloading and reloading the knowledge base. Because the CYC knowledge base is expected to grow to a size of 40 gigabytes by 1994, a more efficient method of preserving integrity is needed. Such considerations illustrate the need for database management methodology in AI systems.

Databases are at the heart of most real-world knowledge base systems.

The third session was chaired by Oris Friesen. In *Query Optimization for Knowledge Bases,* Soumitra Dutta of the University of California, Berkeley proposed a realistic method of calculating the anticipated total cost of query execution by factoring in the cost of optimization. Large knowledge bases require sophisticated optimization techniques, such as semantic query optimization, but including such techniques can lead to prohibitive

. . . the database and AI communities need to come to terms with each other and learn to work together without imposing unrealistic requirements . . . on each other.

optimization costs. Hence, it is important to know when one has reached the point of diminishing returns within the optimization process.

In *Front-End Analysis for Information Knowledge System Design,* Rob Rucker of Arizona State University focused on the need to analyze semantics above the level of conceptual graphs. He proposed a communication schema of some 90 relations to capture semantics that are often lost in data-modeling schemes such as entity-relationship models, IDEF models, semantic nets, and logic diagrams. The audience expressed some interest in this approach as a means of capturing the intent of queries. Considerable discussion took place about whether this approach enhances or duplicates John Sowa's work on conceptual graphs.

In *Object-Oriented Database Design Using Relational Database Methodology,* Yee-wei Huang of Kansas State University proposed using the formalism of relational database methodology to impose some method on the process of object-oriented database design, which tends to be rather ad hoc. A modified form of SQL, extended to support objects, was also proposed. The audience questioned the capability of the relational model to accommodate the semantic richness of the object-oriented techniques.

The last session was a panel chaired by Sara Graves on the viability of database management in AI systems. The panelists were Michel Pilote, an independent consultant; Tony Pizzarello of Bull HN; Forouzan Golshani; and Michael Brodie.

Pilote described a real-world experience in developing an expert system application for insurance underwriting. One of the lessons learned from this effort was that it is important to capture not only the expertise of the application specialist but also that of the data processing professionals. Another insight was that designing and coding from scratch seems to be advisable rather than forcing existing database packages or expert system shells onto the application solution (it especially facilitates technology transfer).

Pizzarello noted that one area where AI can assist database management is in the development of new applications which use data that already exist on large systems. However, many applications already exist for which AI is of little use. One such application class is the airline reservation system, which has always focused on performance. Another such class is the banking system, which has focused on performance and data integrity but has essentially ignored the technology of relational databases. The last class is the manufacturing system, which has always had an overriding concern for data integrity. It seems that AI has the least to offer this last application class because it is too indeterminate and heuristic.

Golshani observed that from a mathematical point of view, knowledge-based systems and database systems can be seen as similar, but obviously each has its own special characteristics. For databases and expert systems to be successfully integrated, it is essential to first find a formal framework that can adequately capture the characteristics of both systems. Having defined all the necessary capabilities of the integrated system in a single conceptual framework, one then seeks appropriate technologies to meet the efficiency requirements. Finally, Golshani reported on an expert database system project currently under way at Arizona State University.

Brodie began by pointing out that heterogeneity is a fact with which we must deal. We cannot define a new and best ontological model to be used by everyone (as some AI people have proposed). Therefore, the database and AI communities need to come to terms with each other and learn to work together without imposing unrealistic requirements (such as starting with a clean slate) on each other. He also predicted that the principal DBMS vendors will soon become the major AI vendors, in part because the vast majority of rules used in real-world applications can be accommodated by a relatively simple inference engine (for example, the XCON/R1 expert system). Based on observations gleaned from his attendance at a previous AAAI workshop on the functional requirements of large AI systems, he noted that most AI researchers are focusing on issues the database community wrestled with during the 1970s and early 1980s. One notable exception is Jaime Carbonell's project at Carnegie-Mellon that focuses on large-system issues.

Commentary

A selective list follows, in relatively random order, of some of the more interesting and provocative comments, questions, and discussions overheard throughout the day. (This list is intended to convey some sense of the diverse, sometimes conflicting dialogue that occurred during the workshop.)

> With the use of database management techniques, coupled with knowledge base approaches, we have improved our productivity tenfold in entering and manipulating images.
>
> Most of what is being done in AI has been done by programmers for a long time, but AI is failing to show how it can be used to make problem solving easier. It seems that some AI knowledge base researchers are rediscovering programming rules long known to application programmers.
>
> AI should not be allowed to be used to promote projects such as DARPA's Pilot's Associate Program, for moral reasons.
>
> AI researchers tend to ignore work that has already been done by database researchers (for example, data-integrity enforcement, concurrency control).
>
> What has been done by the database community is of little direct value to AI because existing programs are not general enough, and they are difficult to extend.
>
> AI rules are inadequate to deal with many real-world application problems (for example, airline reservation systems).
>
> The rules needed in application programs are computationally very complex.
>
> Most rules needed in application programs are not very complex conceptually (for example, such applications generally don't require recursion, theorem proving).
>
> It is sometimes easier to interface a KBS to relatively unstructured COBOL-generated output than to a database.
>
> The relational model seems inappropriate to capture the extensive semantics of the object-oriented model.
>
> Perhaps linguistic analysis can be

mapped onto a set of core relations to provide a semantically rich mechanism that interfaces to numerous natural languages.

Although AI systems need to recognize and deal with degrees of inconsistency, they must also be capable of guaranteeing data integrity and consistency at the database level. This is a challenge for AI.

Can knowledge bases accommodate a distributed database environment where there is no central knowledge repository, such as with a network of powerful workstations communicating on a peer-to-peer level?

How will KBSs perform when forced to deal with millions of frames?

Security in KBSs has not been dealt with very extensively.

Concluding Remarks

The workshop was successful on two levels. It demonstrated that a great deal of miscommunication and misunderstanding between the AI and database worlds still exists. Also, it made at least a small contribution toward narrowing this communication gap.

Database technology can be useful to the AI community. However, it is not adequate to merely hand over finished database products and tools to the AI community and expect them to be used. Rather, the database community needs to gain a better understanding of AI requirements so that an informed correlation can be made between what is needed and what can be utilized to fulfill these needs. (For example, traditional transaction management concepts are needed in AI, but they need to be enhanced with support for long transactions as well.) Only then is it possible to build on the body of knowledge that exists in the database realm (for example, in the areas of concurrency control, data models, and transaction management) and eliminate redundant research and development in AI.

Such an effort demands that database people become more knowledgeable about AI requirements and that AI people become more informed about what has been achieved in the realm of database management. To do otherwise will lead to yet further reinventions of many different wheels.

Oris D. Friesen is an Advanced Technology Engineering Fellow in systems support software for Bull HN Information Systems, P.O. Box 8000, Phoenix, AZ 85066. He is responsible for the insertion of new technology in the areas of software engineering, AI, and database management. His interests include databases, knowledge bases, human-computer interfaces, manufacturing automation, and technology management. He holds B.S. and M.S. degrees from the University of Arizona and a Ph.D. from Arizona State University.

Forouzan Golshani is the director of the Intelligent Information Systems Group and a faculty member in the Computer Science Department at Arizona State University, Tempe, AZ 85287-5406. Previously, he was in the Department of Computing at the Imperial College in London. His research interests include AI, databases, software engineering, and languages. He received his B.S. from the Arya Mehr University of Technology in Tehran and his M.S. and Ph.D. degrees from the University of Warwick in England.