

Intelligent Databases: a program for research and development

Fern B. Halper¹

Miklos A. Vasarhelyi²

October 1991

The authors are grateful for comments received at the International Conference of Knowledge-Based Systems and Classification in Reisenberg, Germany, at comments from colleagues in seminars at Rutgers University and AT&T Bell Laboratories as well as many other presentations of this paper.

¹ **AT&T** Bell Laboratories.

² Rutgers University, Graduate School of Management and **AT&T** Bell Laboratories.

INTRODUCTION

This paper deals with and proposes a program of research on database intelligence linking several areas of more traditional computer science research with management information systems. Parsaye et al.³ defined intelligent databases as *"databases that manage information in a natural way, making that information easy to store, access and use."* They defined three levels of database intelligence:

- i. intelligence of the high-level tools
- ii. intelligence at the user-interface level
- iii. intelligence of the underlying database level

This paper focuses on the second and third levels bringing in the prospective of not only computer tools and single site needs but the more general view and prospective of management information systems. Today's macro corporate and governmental databases incorporate substantial level of detail and history typically at the event level of digitally recorded information. These incorporate textual and numerical information about the main entities represented, its attributes and relationships. In the world of data processing, most companies/industries will have a major data storage need and focus specially related to its main end-activity. For example, the largest databases of phone companies relate to individual phone calls, the largest databases of the IRS deal with taxpaying entities (individuals and companies) and the large databases of insurance companies contain insurance policies and their attributes. In addition to these main activity-related databases most entities will have complex financial and production oriented systems. Despite last decade's myth of the "Integrated Information System" the reality is that only small to medium systems can be integrated and contained in one single data processing entity or into a cluster of closely coordinated devices.

Database technology has evolved in terms of key paradigm from the hierarchical and network models to a current preference of the relational model basically for the support of analytical functions. The reality is, however, that currently large corporate systems continue using the hierarchical model while commercial software gears up to support relational databases of substantive size. Performance problems substantially limit the potential of current relational software when the context goes into the multiple gigabyte or terrabyte domain. Furthermore, distributed file system technology is much more prevalent at

³ Parsaye, K., Chignell, M., Khoshafian, S. and Wong, H., "Intelligent Databases: Object-Oriented, Deductive Hypermedia Technologies," New York, John Wiley & Sons, 1989.

the workstation and PC-net level than at the mainframe or mixed network architecture levels.

Furthermore, an emerging trend is calling for the creation of multimedia, event and structure oriented **"object databases"**^{4 5 6}. These future databases in addition to possessing many of the characteristics of extant databases will substantially expand media and attributes of the elements being manipulated.⁷

While corporate databases are still viewed as retainers of traditional information as described above, personal computers, and voice systems (respectively containing pixel images and sound structures) now retain a set of information that eventually will be considered as integral part of the corporate databases. These two additions to the traditional binary code descriptions of text and numbers expand dimensionally the scope and problem of corporate databases. If film and sound libraries (a more complex set of the above elements) are considered and anticipated to need similar linkage, addressability and processability as current data processing elements the problem expands even further.

The area of imaging will further compound the problem as it will add in to traditional data processing the entire domain of paper document retention and its inherent problems. Advances in scanning and OCR technologies make these events a very likely near-future development strongly driven by market demand in terms on need for productivity improvement and an ever increasing need for sophistication of the analytical information set.

Figure 1 illustrates the main elements of data for future databases. Extant technology addresses issues of translation and conversion on focused often ad hoc basis. Voice synthesis and recognition relate voice sounds to the magnetic image of its component letters. OCR converts pixel images to the magnetic representation of identified letters. Printing converts magnetic representation to visual (printed) images and so on. No overall structure of

⁴ Jordan, David, "Understanding Object-Oriented: A Unifying Paradigm," *Communications of the ACM*, September 1990, pp.40-60.

⁵ Zdonik, S. "Object Management Systems Concepts," *Proceedings of the Conference on Office Information Systems*, ACM/SIGOA, Toronto:1984.

⁶ Butterworth, P., Otis, A., Stein, J., "The GemStone Object Database Management System," *Communications of the ACM*, Vol. 34, No. 10, October 1991, pp. 64-77.

⁷ Silberchatz, A., Stonebraker, M., Ullman, J., "Database Systems: Achievements and Opportunities," *Communications of the ACM*, Vol. 34, No. 10, October 1991, pp. 111-120.

representation and translation among different media exists or is expected to exist in the nearby future.

In recognition to the problems of compatibility among media described in Figure 1 or even the graver problem of relating entities of very different nature some advance has occurred with the advent of object oriented databases¹ and the work of defining and performing operations on objects of non-mathematical nature.

Current database technology is in a preparadigmatic stage. Considerable part of today's research effort still focuses on the development of storage medium (increasing the density of magnetic storage media), the development of more efficient relational models, and efforts of linking larger magnetic and/or optical storage media.

It is clear from the above discussion that linear or even exponential expansion of magnetic storage technology will not resolve the plethora of problems and expanding needs that has already appeared. Consequently, deterministic methods and structural artifices will have to give way to a superior database model. This paper focuses on the concept of developing **intelligent databases** drawn from the human information processing model to try to tackle data retention problems that are arising. The concept of intelligent databases does not imply a data structural model but a family of solutions impounding intelligence in the different elements of the process as well as making its main elements interact in a functional and stochastic manner.

The first part of this paper introduces and motivates the paper, the next section defines the concept of **intelligent databases**, the third section of the paper focuses on the current model of data processing and on issues for intelligent database design while the last section proposes a plan for research and identifies the key problems to be resolved.

INTELLIGENT DATABASES

Parsaye et al. base their intelligent database model on five information technologies: (1) databases, (2) object-oriented programming, (3) expert systems, (4) hypermedia and (5) text management. This approach is useful in the construction of axioms for intelligent databases but rather restricted in the ability to postulate a program for research in databases.

¹ Zdonik, S. Op.cit.

There is little reason to believe the human information processing model (**HIPM**) to be the ultimate in intelligence and storage. However, there is no question that it is superior in multiple features to the data processing model (**DPM**). Consequently, it forms a proper basis for axiomatic comparisons in this paper^{2 3}. Table 1 introduces a comparison of features and some evaluative comments:

Table 1

	HIPM	DPM
Retention	Gradual Erasing Gracefully degrading	Binary Permanent Binary
Structure	Event Oriented Associative	Event Oriented
Medium	Neural, chemical	Magnetic/optic binary
Processing Mode	Parallel, distributed	Sequential

Retention

Human memory has been often classified into three categories: short term, medium term and long term. A large set of events is recorded by the sensing instruments (vision, hearing and touch) and used for immediate guidance. Part of this sensed information is automatically ignored while a subset is used for immediate purposes like balancing of steps in a walk, hand control when grabbing an object, etc., frames of visual memory are kept to relate to sequential events. This immediate/short term memory is filtered for medium/long term retention. Cognitive processes, which are highly structured meta-processes may substantially affect the retention and allocation of memory frames.

Studies using neural images of word reading related to the more classic approach of studies on patients with brain

² Human machine comparisons have been extensively used in the literature. Von Newman, one of the early designers of computers in his posthumously delivered speech, "The computer and the brain" demonstrated that insight could be gathered in such a manner.

³ James, W., "Dumb Intelligence," *Electronics World+ Wireless World*, March 1990, pp.194-199.

lesions⁴ have improved the comprehension of the location of certain cognitive operations in the brain. These are linked to theories and models of brain processes in hypotheses about ways of thinking such as the associative model⁵ leading to philosophies of computing such as the current neural network⁶ approaches that use different algorithms within a generic theory.

Application databases typically receive data from one single type of input device. The nature of the data is digital and typically ASCII (or EBCDIC) in representation. Data retention is complete with no context dependent data retention or filtering mechanism. Retention of data in the DPM is controlled through fix time policies and little context dependency except at a macro level. Portions of the data are stored at different access levels some of it available in main memory, other in direct access devices and a large portion in sequential files requiring manual intervention for access.

Structure

Current corporate databases use two main criteria for structural organization: organizational/application structures and data processing facility structure. The first organization/application criterion typically looks at the data focusing at three main categories: (1) **organization**, (2) **responsibility** and (3) **expense** (revenue) codes. The second, are contingent mainly at the way the company's computer facilities are organized for example one large centralized data processing facility, several regional data centers or applications distributed over the country⁷. Research has suggested alternative approaches for logical organization of accounting structures⁸ but practice has not yet implemented or tested these approaches. An event oriented data organization, if such an approach can be well implemented, may present a more natural environment for data storage in an HIPM like processing model.

The HIPM can suggest alternative ways of structuring data as well as provide additional insights into data storage intelligence. The human brain seems to work as a set of parallel processors working in large clusters located in the

⁴ Posner, M.I., Petersen, S.E., Fox, P.T., & Raichle, M.E., "Localization of Cognitive Operations in the Human Brain," *Science*, Vol. 240, June 1988, 1627-1631.

⁵ Hirai, Y., "A Model of Human Associative Processor," *Transactions on Systems, Man, and Cybernetics*, Vol. SMC-13, No. 5, September-October, 1983, pp. 851-857.

⁶ Caudill, M., "Neural Network Primer," *AI Expert*, 1990.

⁷ A large NY bank has its credit card operations located in the South, its consumer credit operations in the Midwest and demand checking in NY city. The linkages of these systems, a marketing, must be cumbersome and expensive.

⁸ McCarthy (1979, 1982) has proposed an "events" oriented approach focusing on the events that occur in an accounting environment.

different areas of the brain and being helped by some degree of independent device control from the different parts of the body.

Current data processing technology is evolving towards multi-processed machines as well as chips with many processors imprinted on it. The phenomenon of multiplication of processors in a single hub generates the need for parallel processing oriented software and consequently parallel processing oriented database management and access.

The second consideration is the evolution of the distributed processing technology whereby distributed file systems are automatically managed by a network control software and protocols. In this processing architecture, often composed of low-cost workstations and backboned by a high transfer rate local area network, intelligence about storage content and processing capabilities allows for improved utilization of resources.

Medium

The DPM uses primarily binary recording on magnetic medium now expanded by optical, still digital, means. The expanded family of corporate records described in Figure 1 expand the nature of records particularly encompass non-processable analogic images and incompatible voice processing.

Neuro-physiologists still do not understand well the medium, storage and processing of the brain. It is not clear whether memory and processing are intermingled in the HIPM or there

is specialization of functions separating storage, processing and/or control. There seems to be some evidence that there is information on the chemical medium of the brain, that

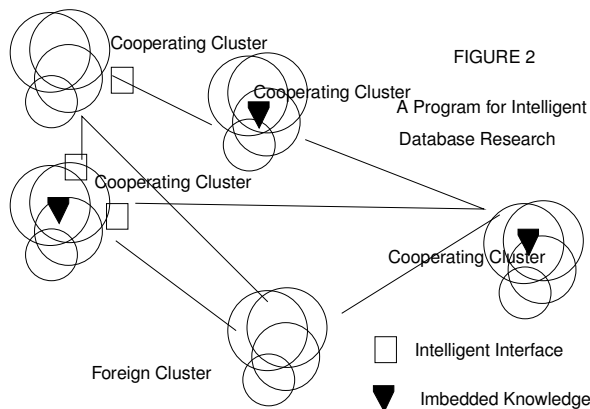


FIGURE 2
A Program for Intelligent
Database Research

synapses are positioned and link neurons tailoring thought processes and analogic information structures as well as some possibility the information is imprinted into the DNA structures.

Databases currently present very rigid and unforgiving media and structures. Soft organization, with substantial influence of knowledge structures may be of great value to improve processes from purely deterministic to more representational and similar to the superior (if not in all dimensions) HIPM.

Processing Mode

The understanding of human data processing is also of great use for developing the issues related to pattern identification in database intelligence. Our understanding, of these processes, despite great advances in the last decade is still sparse, however questions such as the ones stated next posit the need for focus on a different set of processing issues, typically more of stochastic and knowledge nature.

When should attention be paid?

Deals with priorities and interrupts in the dealing with the collection of data in a constant stream.

How to cope with unexpected events?

Concerns the reinterpretation of wired-in models or their incorporation into existing structures.

How to learn without a teacher?

Indigenous knowledge acquisition structures.

How to select a combination of facts that is relevant for a particular situation from one with irrelevant facts?

Filtering and model fitting.

What are the processes to rapidly identify familiar facts in a sea of data?

Fast prototyping and general feature identification

How to combine knowledge about the external world with information about the internal world (needs, structures) in order to satisfy system objectives?

Coherence of knowledge structures.

These above questions clearly illustrate the major differences between the DPM and HIPM and particularly the need for soft, knowledge based, information processing functions. On the other hand the responses to the questions strangely use terms and concepts from extant DP technology.

Next section examines some emerging issues in this DP technology.

DATA PROCESSING

First it is desirable to examine the future and what is already in the horizon of applications or in the emerging research & development literature. On a macro level the major immediate technological developments entail⁹: workstations, bulk telecommunications, mass storage and expert systems. In an intermediate period the development of optical computing, neural network computing and the long term potential of organically grown genetically engineered computing presents great potential for less primitive computing devices and closer resemblance to the HIPM.

Cooperating Computing

Of great potential is the concept of cooperating computing whereby a corporate MIS or net of computing devices cooperates not only on performing requested tasks but on participatory management and on the development of knowledge about themselves and on the distribution of tasks and specialization. The issues of cost chargeout and allocation are an obstacle for the rationalization and distribution of power. New algorithms and approaches to the allocation and distribution of telecommunication and data processing costing (and transfer pricing are necessary for successful commercial implementation of this approach.

Current work focuses on the distribution of jobs¹⁰ between clustered processors, typically concerning the same machine and multiples CPUs as well as on the concept of distributed file systems whereby information about storage device content is shared either through constant updates or through a rigid protocol of addressing.

In terms of cooperation what is needed is protocols that handout of processes when processors identify themselves as busy while receiving signals that others are less occupied or more adequate for particular tasks. Learning about the

⁹ Vasarhelyi, M.A., and D.C Yang, "Technological Change and Management Information Systems," *Proceedings of the Twenty-First Annual Hawaii International Conference on System Sciences*, Hawaii 1988, pp. 191-197.

¹⁰ Merges, M.J., Benhenni, R., Drakopoulos, E. & Hariri, S., "Design Guidelines for Distributed Computing Systems," *Private Communication, AT&T Bell Laboratories*, April 1990.

nature of their job-mix, insight into their own capabilities and status of repair (or disrepair) and the ability to look ahead of processing needs and take cooperative action.

Furthermore, cooperative computing needs to change its focus from merely participatory to proactive and self-insight oriented meaning that idle time be eliminated by being dedicated to constant reevaluation of its own structure and capabilities, of the structure and capabilities of its peer (cooperating) group. The issues related to representing processing power, estimating processing needs, determining frequency of communication, nature and volume of information handoff and sharing are of great import and barely touched in the literature.

Of great importance in research about cooperation is the concept of concurrency control¹¹ whereby management of semi-simultaneous access is performed through the operating system. In distributed or/and cooperating systems the time scope of concurrency is exacerbated and must be resolved gracefully to avoid major deterioration of functionality.

Despite the progressive blurring of the concepts of storage and processing that will occur it is worthwhile to focus on a sub-item of cooperative computing that deals with distributed or cooperating databases.

Distributed Databases

Cooperation among processors and facilities can be achieved in many forms. Let us assume that architecturally we will call any cluster of CPUs that are physically connected by a common hardware BUS as opposed to some form of local or wide area network is called a machine. Several machines interconnected by what is currently called a LAN are a cluster and a network of clusters dedicated to substantive cooperating (as opposed to just communicating) are called the **network** of cooperating computers.

The concept of **data warehouse** has emerged in many MIS applications however from the current standpoint it is only another cluster dedicated mainly for storage. Furthermore, cluster gateways are also ignorable from the standpoint of data storage and retrieval as they are exclusively communication access devices.

The introduction of this paper clearly showed that two main phenomena are happening: (1) a greatly expanded scope of data storage needs and (2) a major change in the nature of

¹¹ Barghouti, N.S. & Kaiser, G.E., "Concurrency Control in Advanced Databases," *ACM Computing Surveys*, Vol. 23, No. 3, September 1991, pp. 269-318.

the data to be stored progressively moving away from pure digital representation to a future of enriched analog representations not yet envisageable in its full scope.

Medium interchange processes are still in the early stages of research but progressing to a point where for example the interchangeability of paper and magnetic text are progressively of greater ease. The same is not true for images or sound which cannot easily or effectively be converted into ASCII and operated upon. The inclusion of image and/or sound into documents is typically a segmented process with foreign object insertion characteristics.

The bases for current data distribution typically relate to storage limitations and organizational scope. Interorganizational cooperation and distribution are far in the future. On the other hand traditional issues such as data redundancy, backup and access have been studied and are constantly in the mind of developers.

The major need is of introspection in database operations. The intrinsic nature of the growth in databases makes traditional search, indexing and keying cumbersome, expensive and sometimes close to unfeasible. In the first blush it is necessary to create three levels of intelligence in databases in addition to the current networking capabilities of some relational products: (1) ability to understand its own content even if only at a primitive level, (2) ability to dynamically rebalance itself among storage access media and clusters and (3) intelligence and sharing of insight among cooperating storage/processing devices.

Content understanding entails the ability to see how its main content (entities, attributes and schema) is used and interrelates. At a primitive level prediction of frequency of access, timing of queries, interrelationships of queries, nature and volume of updates as well as purging issues.

Dynamic rebalancing entails the ability of the database to **decide** on bringing information in and out of faster access media from slower (mass storage) media and over time obsoleting the notion of purging to a notion of progressive degradation of access. Furthermore, still on a primitive level entails the usage of its understanding of indices and index usage for moving information from cluster to cluster and the creation of additional indices or "guessing schema" on the location of information.

Insight sharing entails the progressive usage of not primarily driven processor time to gather insights on different cluster databases and sharing of these insights.

Database Content Representation

In order to be able to understand and represent itself, databases must have forms of representation of its knowledge, content, use pattern and dynamic characteristics. Presently most large and middle sized systems have extensive accounting tracking and statistics on time, usage and customer. The same information also can be obtained at a more data specific level from DBMSs. Over the years a series of products that statistically manipulate these accounting traces have appeared in the market. The next generation of products and research must be able to proactively use usage representation for rebalancing and self activate in the search for self-understanding.

Database Structure

While current commercial DBMSs tend to present a main form of structural models (e.g. hierarchical or relational) future DBMSs must deal with a multiplicity of models and a multiplicity of object types. The next generation of databases must, for example, gracefully be able to attach pictures of group members to an organizational graph and to extensive accounting information on the organization. Once this flexibility is achieved the progressive integration of analog information to digital files as well as the adoption of alternate database models (e.g. associate, adaptive) will be facilitated.

Dynamic learning about progressive database structures must occur automatically in a database leading to issues in many areas such as: representation of events, representation of attributes of events, representation of entities, operations with entities, ability to dynamically define and redefine cluster of entities, ability to dynamically attach and detach attributes to entities and clusters, ability to create and differentiate between soft and hard links among data, ability to create classification schemes and to improve automatically these schemes, etc.

The above issues probably lead to a series of more pragmatic issues that seriously hamper today's operations on large databases. These issues are issues related to the content of these large databases. The limitation of knowledge domain or function can lead to promising applications of artificial intelligence within a shorter time-frame. For example, O'Leary¹² focuses on accounting systems and proposes the use of daemons and objects as "useful devices to facilitate the

¹² O'Leary, D.E, "Artificial Intelligence and Expert Systems in Accounting Databases: Survey and Extensions," *Expert Systems with Applications*," Vol. 2, 1991.

organization, storage and application of intelligence for accounting database systems."

Database Content

Two issues of very practical nature impinge on the wider set of knowledge issues raised earlier in this paper: (1) pattern matching and (2) automatic indexation. These problems can be examined from the context of a wide array of knowledge and expert system issues to a practical view of what happens in large databases. If some of the practical problems can be resolved major steps in setting the foundation of intelligent databases will have been performed.

Advanced pattern matching: corporate databases are composed typically of a set of independent applications using at best common data indexed in alternate scheme. If some degree of integration has been achieved controlled redundancy exists and is a major step for the planned performance of application functions. the problem however of duplication (logical) for related entities is a major one. Telephone companies suffer from the problem of identifying individual clients that moved and/or changed names and cannot be tracked for marketing purposes. Accounting systems suffer from vendors, clients and third parties that are the same but cannot be identified as such, banks suffer from services being delivered to the same entity but the entity not being identified. To this problem extensive mechanical manipulation has been applied achieving mixed results. It is impossible at the current stage of research to identify logical duplication without a particular search for it.

This area has extensive practical implications and must evolve from mechanical matching algorithms to knowledge based matches as well as the extensive usage of foreign information bases to help in the development of information profiles.

The problem of **automatic indexation** occurs in many areas of research but is of special importance in this context. In order to understand the meaning of a particular stream of text frames ¹³ must be built to explain and delimit the domain. A large set of information is today text oriented in the business world such as information about the competition, documentation of products, financial statements, customer complaints, etc. Of great essence is the development of ability of crude understanding of text for primitive classificatory purposes.

¹³ Schank, R.C. & Childers, P.G., *The Cognitive Computer*, Addison Wesley, 1984.

A Plan of Experimentation

In addition to vague textual understanding and advanced pattern matching experiments in the development of primitive cooperating environments must be performed. The following steps propose a first approach to such a work within the UNIX environment.

The UNIX operating system is by its intrinsic nature ideal for the development of an experimentation plan. This plan aims to prove the feasibility of sparse understanding of external database content (as opposed to full data dictionarization as in an NFS approach) and repeated self reorganization based on a series of experimental-based or knowledge-based algorithms.

A>>> define a set of cooperating workstations

B>>> add local disks to their configuration to contain the experimental databases

C>>> install the COSHELL¹⁴ as a local network coprocess server to allow for the distribution of processes among cooperating workstations on top of the Korn shell¹⁵.

D>>> trace, accesses and maintenance routines that are frequent and/or prescribed in the original database (that was distributed for experimental processes).

E>>> devise an experimental algorithm that will have maximum likelihood of finding data in the first local search. Migrate process from processor to processor finding data clusters.

F>>> record search paths and improve on-these based on rules to be recorded in cooperating device.

G>>> broadcast key aspects of the experience to all operating systems of cooperators or more likely post a representation of this experience in a common polling locus.

H>>> fire off daemons at usage and timing thresholds for self-organization.

I>>> run experiments on: self-reindexing, methods of process migration, rules of process migration, elements of key representation.

The needs and features of research towards intelligent databases are multiple to the point that some logic is need in the establishment of priorities and the practicality of

¹⁴ Kintala, C.M.R., "COSHELL," *AT&T Bell Laboratories, Internal Communication*, September 20, 1991.

¹⁵ Korn, D. "the K-Shell," Addison Wesley, 1991.

where to proceed. The concluding section of this paper addresses these priorities.

CONCLUSIONS AND A PROGRAM FOR RESEARCH

This paper discussed issues on database intelligence with an emphasis on comparing the DPM with the HIPM. The traditional analysis related to computer science issues was enriched with corporate data processing and MIS considerations.

An **intelligent data system** entail a network of **data processing cluster** sharing information on their **internal content and structure** and dynamically and automatically updating this knowledge **acting upon this knowledge**.

Figure 2 describes such a system whereby some form of network intelligence exists in the management of independent cluster resources, the flow of data among its elements and the reallocation of data and processing resources.

A program of Research

The philosophy of a program for intelligent database research may revolve among two basic concepts of database knowledge: (1) introspective knowledge and (2) extrospective knowledge.

Introspective knowledge entails studies of what the database should know about itself: its structure, its content, its frequency of update, the nature of the queries, volume of data in its segments, the relationships among its parts, the nature of the content the security and privacy considerations in its usage and volume of usage patterns across time and users.

Despite extensive studying of database structures, models and implementation both academia and practice has been woefully in the deep understanding of ways of describing database structure at the meta-level, database usage, nature of data content, etc. This understanding, its parametrization, its metrics and distributions are essential for the development of the natural next step of research where databases use this knowledge for self-organization purposes.

We must be able to develop DBMSs that: identify when they are idle or have low usage, can trigger a self evaluation resulting on representational structures placed in storage, can fire off actual self-reorganization in particular relating to physical location, indexation, and data retention.

Extrospective knowledge entails the development of representational protocols that can be exchanged among hosts (clusters) giving a view of its environment and setting the rules for acceptable cooperation. Modern research on networks, and their practical implementation, led to a substantive approximation among local and foreign clusters where by the first are part of the extended family while the second only allow and promise very limited (and highly controlled) cooperation.

This understanding without full context, its parametrization, its metrics and distributions are essential for the development of the second natural next step of research where databases use this knowledge for authentic cooperation and resource sharing.

In conclusion, it seems that the comparison of the DPM and HIPM lead to some insight on the development of intelligent databases. The understanding of the HIPM is still very incipient and the linkage between the human neuro-physiological level, the logical thinking processes, the control of physical motion and human emotions is close to nill. Stridns in the understanding of the HIPM should also give us ideas on improving our DPM. It must be remembered, however, that the hardware and objectives of these two models are different and they will ultimately each have its strengths and weaknesses in comparative terms.

Of immediate need is work on two bases of self-insight for databases: (1) advanced pattern matching and (2) self-indexation.

Of basic research nature we find extensive need for the development of knowledge representation algorithms and structures of its utilization for learning. Learning algorithms and philosophies must cover a wide range. Learning about a structured phenomenon where examples and substantial data exists, learning about the changing nature of environments, learning about issues that must be learned about and learning about learning.

Two main locations of knowledge seem to be obvious in the discussions within this paper. First knowledge at the hub (gateway) of a cluster about traffic, data processing status and state of the network. Second, boxes of introspective knowledge in the content of the database, constantly being affected by database activity.

The issues of cooperative data processing and cooperating databases cannot be separated leading therefore to the need of substantive research on cooperating processors and consequently on one of the major stumbling blocks which is non-sequential (parallel, hierarchical) software.

REFERENCES

- Alter, S. , *Decision Support Systems: Current Practice and Continuing Challenges*, Addison-Wesley Publishing Company, 1980.
- Baeza-Yates,R.A., "Improved String searching," *Software Practice and Experience*, Vol.19, no. 3, March 1989, pp.257-271.
- Bailey, A.D., Graham, L.E. and Hansen, J.V., "Technological Development and EDP" in Abdel-Khalik, A. R. & Solomon, I. "Research Opportunities in Auditing: The Second Decade," American Accounting Association: Auditing Section, Sarasota, Florida, 1989.
- Bailey, A.D., K. Hackenbrack, P. De and J. Dillard, "Artificial Intelligence, Cognitive Science and Computational Modeling in Auditing Research: A Research Approach," *Journal of Information Systems*, 1987.
- Buchanan, B.G. and E.H. Shortliffe, *Rule-Based Expert Systems*, Addison-Wesley Publishing Company, 1984.
- Carpenter,G.A. & Grossberg,S., "The art of Adaptive pattern Recognition by a Self-Organizing Neural Network," *Computer*, Vol.21, March 1988, pp.77-88.
- Chapnick, P., "Intelligent Databases," *AI Expert*, March 1990, pp.5-6.
- Data Processing, "Intelligent Database," Vol. 27, May 1985, pp.50.
- Fox, C. and F. Zappert, "Information Systems Evolution in the Near Future," *AT&T Bell Laboratories Private Communication*, Holmdel, NJ. December 1985.
- Hendrix, G.G., "The Intelligent Assistant," *Byte*, December 1987, pp. 251-258.
- Kahan, S., T. Pavlidis, and H.S. Baird, "On the Recognition of Printed Characters of any Font Size," *AT&T Bell Laboratories, Private Communication* (January 1986).
- Kelly, K., G. Ribar, and J. Willingham, "Interim Report on the Development of an Expert System for the Auditor's Loan Loss Evaluation," *Artificial Intelligence in Accounting and Auditing*, Markus Wiener Publishing Company, 1988.

Kiessling, W., "Access Path Selection in Databases with Intelligent Disk Subsystems," *The Computer Journal*, Vol.31, No.1, February 1988, pp.41-50.

Lippman, R.P., "Pattern Classification Using Neural Networks," *IEEE Communications Magazine*, Vol.27, November 1989, pp.47-50.

McCarthy, W., "An Entity-Relationship View of Accounting Models," *The Accounting Review*, 1979.

McCarthy, W., "The REA Accounting Model: A Generalized Framework for Accounting in a Shaded Environment," *The Accounting Review*, 1982.

Moss, C.D., "Intelligent Databases," *Byte*, January 1987, pp.97-99.

O'Leary, D.E., "Artificial Intelligence and Expert Systems in Accounting Databases: Survey and Extensions," *Expert Systems With Applications*, Vol. 2, 1991.

Telfer, B. & Casasent, D.P., "Ho-Kashyap optical associative processors," *Applied Optics*, Vol. 29, No. 8, 10 March 1990, pp.1191-1202.

Vasarhelyi, M.A., "Expert Systems in Accounting and Auditing," *Artificial Intelligence in Accounting and Auditing*, Markus Wiener Publishing Company, 1988.

Wolitzky, J.I., "A Low-Cost UNIX Speech Workstation," *AT&T Bell Laboratories Private Communication*, Murray Hill, N.J. July 1985.