On Components and Design of Industrial Information Systems

L. H. Aslanian

Laboratory of Discrete Modeling Technologies Institute for Informatics and Automation Problems of NAS of RA and YSU

Abstract

In this paper the Industrial Management problem as the set of Organizational and Information Securing Activities is studied. Acting in accordance with the Information Standard of an organization the Industrial Management process consists of a Distributed Data Base system, based technologically on the advanced tools and methods of Theoretical Computer Science, Discrete Modelling and Software Technologies. The description of an application activity on design and implementation of Industrial Information system oriented to the case of NIS countries and the transitional period of economy is considered. Although the common technologies of Distributed Data Base systems are well developed, the majority of design projects are faced with the particular design time problems, arising from the special conditions of application areas.

So, during the work naturally arose and were formulated several problems of data retrieving in terms of Pattern Recognition and algorithms of Dynamic Programming. Several theoretical postulations and extensions are considered for these problems as well as the appropriate technological software tools are defined and explored. There were considered the properties of solutions of the Discrete Isoperimetry Problem, created extended structures of partitioning of the set of all vertices of an n-cube related to the computational simple hashing functions, working on the best match searching algorithms. The logical level modelling construction is discussed to create the intelligence shell and software environment providing the Industrial Information and Management problem.

1 Introduction

As soon as the free market economy is in the process of getting actualized in the CCE and NIS, the powerful computer tools for Industrial Management problem are creating a special interest. ¹ The starting realities are that during the Planning Economy period of Socialistic Countries there was a very crude form of planning. Planning was done within the high organizational levels and factories were left to fall on their destiny by using raw materials, technologies and marketing of final production completely. All the results were conditional and there was no real need and possibility of estimation of effectiveness of different parts and stages of industrial processes. There are many other differences of Planning Economy versus

¹This research was sponsored partially by the National Science and Education Ministry of Armenia under grant 96-556, the ArmElectra Inc. and the American University of Armenia.

the so called Western Style, but the main thing is that in order the industry in CCE and NIS to advance it needs to use the special approaches, formalizations and recommendations in form of computer, network and software systems. For the needs of investigation by the different planning strategies the economical consistency and regulation must be in standard and flexibly organized forms. Factories must use and try alternative strategies and select the appropriate one, on the basis of accumulated experience. Therefore, the universal computer aided systems (CAS) of management can be extremely useful.

Industrial information and management infrastructures

The Industrial Information Systems are basically implemented through the local networks with the different types of computers. The whole system is based on the standard operational and data base management facilities. At the same time there are well known particularities of the Industrial Information infrastructures. There is a wide range of information types technological, financial, material and realization support, standards and others. The performance of the whole system must be under the general control of the set of directives, which create the Information Standard of an enterprise. The different constituents of this complicated model are under the permanent reconstruction and change. Therefore, one of the basic conditions to these systems must be the flexibility and the possibility of their reconstruction and extension during the exploitation period. This must be provided by the software system as well as the logical level structures of the information and system models

Some of the basic constituents of general information treatment which will be done within the Industrial Information Management are standard ones. In addition a need of several particular treatment and algorithmical facilities might come up. The implementation must be done in any software environment and the complete algorithmical support of this problems must be provided. Creation of the basic data (knowledge) structures, designing of workstations according to the main partition into the subproblems and providing the system level constructions must be done in any standard data base management environments. Above all this paper there is emphasized the importance of paying special attention to the creation and estimation of different strategies of general management strategies encompassing several ideas.

Therefore, the Industrial information infrastructures including particularities connected with the industrial management problem might be represented on a special professional level

above the standard data base information structures.

Formally it depends of data structures and types of that field, the supposed software environment and the form and consistence of basic functionalities or knowledge forms which we appoint for the considering area of application. If the basic (standard) tools of data base management are well developed and implemented, the particular techniques must be constructed accordingly to the wide areas of applications. One of the main goals of this paper is the description of a system of data and software constructions, which we use to provide the problems of data base environments in a technological level for the needs of industrial management problem. The corresponding applicational software system was implemented in the ArmElektra Inc., one of the typical and powerful factories of the Soviet Union.

If the technological level of a software system is connected with the general flexibility of the software product during the installation and providing, the strategic planning and modelling tools are for solution of the general problem of flexibility in the terms of goals of the implementation field. The main logical level of the performance of a software system

must formalize the tools to configure and maintain the information amount to work out the solution of a set of basic problems of considering implementation field.

2.1 Experimental data treatment on base of pattern recognition

The term Industrial Management is mainly used to denote the process of creation and analvsis of information and decision making for purposes of production planning and scheduling. The process of controlling of components of a production process itself is far wider. It includes the tools of information acquisition by the sets of sensors and the information analysis within the process of production, the testing of quality and the recovering of malfunctioning. The basic structures of experimental data collections in these fields consist of sets of objects. described by the vectors of estimated values of a set of properties of these objects, as well as by the time series, graphical data and images, etc. Particularly, this formats of data correspond to the information of the time of production testing and the process of production controlling. Working with these data the general problem is the recovering of existing regularities in terms of statistics (statistical estimations and hypotheses), pattern recognition (grouping, ordering, learning, prediction, weights of properties), etc. The main difficulty of the implementation of these sets of algorithms is that the recognition methods are heuristic procedures by their nature and their results need to be further interpreted professionally as well as statistically. The exact algorithms like the solution of a quadratic equation gives the exact numbers - solutions. Using algorithms of class of pattern recognition we can obtain a wide range of possible solutions which depends on the idea of the selected algorithm and the badly formalization of particular problems of pattern recognition, which is well known. More exact formalization of these problem moves us far from the application problem which is given only intuitively and on level of a set of classified examples.

Let us consider the case of vectorly described data. Presently, there are many procedures to solve the problem of classification of this data - with stage of learning (classification) and without (cluster analysis). The same approaches and algorithms are recovered for the different applicational fields (biology, linguistics, history, psychology, etc.). The usage of these algorithms in Industrial Management are based on experimental estimations and are not proved theoretically. This is because of the heuristic nature of these problem area. The approximation models are available and they allow us to create the heuristic conclusions only.

During the current stage of development of pattern recognition algorithms several attempts were made to construct the complete mathematical environments (shells) over the classes of heuristic pattern recognition algorithms Zhuravlev [11], Aslanian [2]. This makes the usage of these algorithms to be directed and acceptable. These models include optimization parametric models of algorithms to create the best mode of algorithms in use and current solutions.

The complete solution of the problem of creating of common modelling environments for pattern recognition problems is far from it's end. Perhaps the problem is connected with the fact that the types of input parameters, algorithms of classification and their adaptations are considered in the theory and applications separately. There are a very few approaches of gathering the results and creating common algorithmical modelling environments in the bounds of high intelligent decisions and controls. The idea of this general goal is the comparative analysis of classes and problems of pattern recognition and creation of appropriate algorithmic structures for solution of problems automatically on the basis of properties of

input data and with the estimations of the formulated conclusions. Considering as an example of a local problem of Industrial Management of class of pattern recognition the problem of technical diagnostics on base of sets of sensory estimated values we conclude Aslanian, Karachanian [4], that the pattern recognition algorithms are to be implemented on the base of learning and control sets. Selecting in parametric algorithmic models the optimal values of parameters on stage of learning using the learning information, we have to estimate the level of optimization on the base of the control set of information. As a rule, for the main classes of algorithms - Parametric models of estimation of similarities, Unreduceble tests algorithms and Logical separators algorithms - the selection of parameters might be done also on base of treatment of the learning set by means of Discrete Isoperimetry property.

Best match searching algorithms based on isoperimetry property

The main goal of this paragraph is the description of extended searching constructions, algorithms and software within the advanced Information Systems for seeking the best matches

of information according to the searching conditions.

In practice the described technique was implemented in information systems included in applicational pattern recognition environments as well as in global information systems, in part of securing of works with the Problem Oriented Data Dictionaries. Each application field must supply several groups of terminological and information standards dictionaries, common for the different users and provided, as a rule, through the local network. Dictionaries must be placed on the host computer and the current queries to their entities must be done using the available approximations inputted by users (mistakes in original working documents or entering mistakes, which are supposed to be). The best matches to the input word x must be searched out to choose the exact entry term after that.

The application consistence of these problems are also related to the pattern recognition systems (considered above), supposing the learning stage. Algorithms in this case are trying to find the most similar objects from the learning set F to the considered one - x. On this

basis must be made the final decision about the class or image for object x.

The well known classes of algorithms for seeking the exact and partial matches, as the binary searching trees, inverted file constructions, hashing algorithms and others, are to be implemented in an information system accordingly to the set of local environmental conditions. The algorithms considered in this paper are of class of hashing algorithms based on partitioning of the basic binary space according with the constructions of perfect codes

and/or the Discrete Isoperimetry Problem (DIP) and solutions.

Let F be a finite set of some binary words of length n. x is an input binary n-word. F(x) is the set of all words from F having the (same) minimal possible distance from x(in the simplest case it must be considered the Hamming distance). The optimization of the complexity of algorithms, working with F and x and giving F(x) is to be considered in connection with the special constructions, using through the mapping the basic set F onto the computer memory. The initial methods, based on perfect codes are available Rivest [8]. They are restricted by the very limited set of possible perfect codes. It is well known that the only nontrivial classes of binary perfect codes are Hamming and Golay codes Zinoviev, Leontiev [12] and Tictäväinen, Perko [10].

The DIP problem is one of the typical issues of advanced discrete mathematics. The solutions geometrically are very close to the Hamming spheres. The complete covering of basic binary set by this objects is highly important and must provide more possibilities for construction of searching algorithms. Any positive steps on this direction requires more knowledge of properties of the solutions of the DIP.

So, let us suppose that the basic set R^n of all binary words of length n is divided into the blocks B_1, B_2, \dots, B_m . Accordingly, $L_i = R^n \cap B_i, i = 1, 2, \dots, m$ are lists of elements of F belonging to these blocks and saved as separate lists by h(x) address, common for the elements of defined blocks of R^n . The idea of this searching construction is that using a dynamic programming style algorithm of class of branches and leaves and a partitioning of the space R^n into the geometrically compact blocks we utilise the 0.9% of all input information to get the F(x) by an input x. To achieve this result we have to divide the R^n into the disjoint blocks - solutions of the DIP. So our problem is that neither the splitting of R^n into the solutions of the isoperimetry problem nor the construction of a perfect code for an arbitrary n is possible.

Let a is a nonnegative integer, $a \leq 2^n$. A lexicographic order L_n of the set of vertices of R^n is a linear order, where

- · vertices with more 1's follow the others,
- · vertices with 0 in the first differing position precede the others.

Denote the set of first a vertices of L_n by $L_n(a)$, then by Aslanian [1] $L_n(a)$ is a solution of the DIP. This result was formulated independently by different authors. As it has been recognized in Rivest [8] for blocks B_i , $i=1,2,\cdots,m$ the optimal algorithms of hashing class correspond to the selection of B_i which are the DIP solutions. The extension of these results, giving the description of all of the solutions of DIP also are available. We have:

- The arbitrary solution of the DIP consists of a Hamming sphere of the maximal possible by the given size a radius (denote the radius of these sphere by r_a) and the additional vertices the part of which might be arranged differently.
- For at least 2^{n-1} of cases by a, for n is given the more precise description of solutions of DIP. These sets are included in addition in a sphere, which radius is greater by two than r_a .
- The number of subsets of \mathbb{R}^n with the m interior vertices are to be distributed by the Poisson's distribution with the main value 1/2.
- All the above mentioned descriptions of the considered problem solutions are rather complicated to construct the precise splitting of Rⁿ by this objects. So the geometrical shape of the DIP solutions give us the ideal partitioning objects exemplifying.

There is one more very important item to finalize the step of splitting the basic space \mathbb{R}^n into compact subblocks. That is the usage of different constructions, based on the Hamming spheres of different sizes. As a beginning we can choose a number of form 2^q-1 to be not greater than n. As we know there exists a perfect Hamming code for this case, hence, the exact partitioning of \mathbb{R}^n into the spheres of radius 1. Considering of partitioning of \mathbb{R}^n according with the cartesian product of \mathbb{R}^{2^q-1} and \mathbb{R}^{n-2^q+1} , we can choose the splitting of the first subspace as a Hamming code while the second part might be reminded as is or split

by itself. The main advantage is that the blocks are Hammingean and that the values of the corresponding h(x) function as well as the distances of these blocks from the arbitrary

points $x \in \mathbb{R}^n$ simply calculable.

The further generalization of this idea is correlated with the special representation of arbitrary numbers n by the sums of numbers of form $2^q - 1$. This is for decomposition of \mathbb{R}^n into the subcubes, which can be covered by the disjoint spheres. In parallel there might be entered additional terms for usage of Golay codes and the simple constructions consist of partitioning of arbitrary cubes into the two subspheres. To be short in this part we can formulate the following properties:

• Let us consider a binary vector $\vec{\alpha} = (\alpha_n, \alpha_{n-1}, \dots, \alpha_1)$. The corresponding sum

$$s(\vec{\alpha}) = \sum_{i=1}^{n} \alpha_i (2^{i-1} - 1)$$

is limited by numbers 0 and $2^n - 1 - n$. Moreover, sums of form $\vec{s}(\alpha)$ don't cover these interval completely.

- Sums $s(\vec{a})$ achieve the 2^n-1 different values. The last coordinate α_1 is not essential for the value of $s(\vec{\alpha})$.
- Let us consider the vector $\vec{1}_i$ with the all 1 coordinates on positions $j,j\geq i$ and 0 elsewhere. We'll double the last sum term $-2^{i-1}-1$, which corresponds to the i-th 1 of $\vec{1}_i$. Then we get the numbers from $2^n - 2$ to $2^n - n$ continuously. For the numbers, starting from $2^n - 1 - n$ and smaller we can prove by induction on n, that doubling only one sum term we can receive an arbitrary number from the remainder.

So there are simple properties and constructions providing the splitting of n-cubes into the subcubes, where there exist the Hamming codes and/or constructions of Golay code or trivial partitionings in terms of vertices, spheres of different sizes and subcubes.

Industrial information and management software technologies 3

Traditional information and computing technologies have focussed on well-structured problems involving huge data bases and/or huge numerical computations. Most applications correlated with different application fields often need new algorithmical solutions and high professional level of users is supposed. Different new ideas naturally arise in the during the development of new generation information technologies, correlated with the ideas of creating of technological environments for prototype systems design instead of designing separate systems and creating Intelligent environments to take more and more functions of the users. Those applications have addressed more clever tasks such as those understood pertaining to expert systems, and those supporting decision processes, called decision support systems. Huge national and international projects, including the Japanese fifth generation project and several ESPRIT projects, have supported this interest. From the technical point of view, this has resulted in a move from one machine and user environments to the distributed and integrated environments and the new global networking tools. Artificial Intelligence, Logic programming and Global Networking on base of Distributed Integrated Data Bases exemplifies such application style Mařík, Lažanský [7].

The basic application problem to be considered within the next points is the problem of industrial management in transitional period (distributed integrated information structures for industry with different planning strategies) in CCE and NIS countries. This is one of the so called CRITICAL SOCIAL ISSUES OR EVENTS which is characteristic for the different European countries including CCE and NIS nowadays.

Finally, the theoretical research and recent advance of technology has made possible the construction of integrated distributed information environments and networking for providing information on this basic problems, provide methodologies to allow business objectives to be translated into operational structures, establishing organizational, human and infrastructure resources. Areas of work include tools for business process modelling and description, selecting of relevant information in an environment of information overflow, enterprise systems management, resource reallocation, and simulation of processes and strategies Atre [5], Maier [6], Aslanian, Demetrovics [3], Stonebraker, Aoki, Devine, Litwin, Olson [9].

The development of a powerful and flexible set of tools to support business modelling and decision making is based on creation of problem oriented tools and integration with the operational information systems elements. The global technology of applicational system

will consists of:

- The Extended Database Management System,
- The Problem Oriented Spreadsheet Generator,
- · The Report and View Forms Generator,
- The Functional Subsystems of Administrative Units,
- Special Managerial Utilities and Decision Making,
- System Utilities.

The user oriented front-end interface and network applications must be provided.

Of course, the existing environments (MS-DOS, WINDOWS, UNIX etc.) provide the users with similar universal tools, but many applications require some specific routines that are not considered in standard systems and need to be specially programmed. There exists another approach, which is proposed in this project, where the development of the complete program shell makes possible to design more flexible software components, and the access to the options of primary modules gives opportunities to enhance the product according to the specific requirements of applications.

3.1 The preliminary Description of the Software System

The information systems, based on the database technologies, consider separate available conceptual constructions, based on the theoretical investigations of data models (relational, hierarchical etc.), to ensure an opportunity to deal with large data volumes in real-time interval and in accordance with the standard requirements to the data bases. The data base conceptual level model design on the base of relational data-modelling technologies was carried out within this work.

In accordance to the above mentioned approach, a software system is designed and made out for MS-DOS environment using CA-Clipper 5.3 System. During the last two years a

computer aided system of management for one of the Armenian enterprises was elaborated. It includes the standard data processing tools and in addition to them:

- The Salary Calculation Subsystem,
- The Technological Subsystem,
- Subsystem of Inventory and Product Control,
- · Subsystem of Financial Analysis,
- · Subsystem of Planning Strategies.

The Database Management System realizes the general data control. It ensures the data integrity. It controls the access to the files. The access to the data is organized by the "windows". The windows' attributes are defined by the user externally. Those are the names of the windows, its coordinates on the PC's screen and color, its columns, the fields and their sequence, the related files with the fields of relation, the links with the system data dictionaries etc. The configuration file gives the opportunity to create a "user's view".

The Spreadsheet Generator gives the opportunity to create spreadsheets using as cells the fields of the database file. The user can define expressions consisting of the database

fields simultaneously while defining the window and while working with it.

The Report Generator is an independent unit of the system. It provides with the opportunities to create reports on the database files. The format of the report is defined by the user. It can include the headers, footers, line separators, pagination, horizontal and vertical sums, formulas etc. The user can create reports in three languages (Armenian, English and Russian).

The System Utilities or the service unit of the system contains some standard functions

such as calculator, file manager etc. Others include some specific functions.

The Technological Subsystem and the Subsystem of Planning Strategies are based on the following information structures. The main file consists of the description of Product kinds through the binary inclusions: a technological object (material, detail, operation or product) is included in another technological object. So the Production level is provided and the description tree for the Production types are available from this file. The available intersections of information coming from the different Productions descriptions in a common file are eliminated within this description which minimizes the information sizes. In this case the typical operations of any technological workstation are the definition of a new production, the checking of information on completeness, deleting the information of any obsolete production as well as the estimation of amounts of materials, operations and details for providing of a set products to produce. This is just the point where might be done the decision making and the strategic estimation and selection of the industrial management problem.

The Salary Calculation Subsystem and the Subsystem of Financial Analysis are highly interconnected.

The Subsystem of Inventory and Product Control is the place where the differences of managing in industry in different approaches to the economy, planning and market are emphasized.

It is obvious that the primary copies of the most developed systems need to be redesigned in several aspects during a short period of time. Namely for perfecting different parts and providing the current operational instructions, the network structures and specific managerial functions. We think that nowadays being enriched with methods of management and decision making mechanism, the system can be successfully implemented in any enterprise independently of any economical and of planning rules.

The main objective of the project for ArmElectra Inc. was the development of advanced universal tools for industrial management to formalize the modern methods of management and to include them into the CAS. There are two basic constituents to make the software system technological.

The first is the development of the basic system architecture, where different objects are defined: data base files, index files, relations within the files, view or professional points, working spreadsheets and windows, set of standard operations like total sum checking, functional relations among the fields and groupings of different files, information entering and additional field adjusting strategies, current operational information browsing, creation of report forms, planning strategies, estimation, automation of information interconnection and recovering, as well as subject oriented dictionaries. The main idea is that the basic software system is an empty shell to work with the structures and objects defined above. Entering and changing freely the consistence of this object descriptions we really create and change different applications.

The second component is a standard and well known advantage of many algorithmic languages - the macro programming facilities. In this case the CA Clipper 5.3 is used and full excellence of enhanced programming components are implemented.

Let us consider the consistence of spreadsheets and windows files' as an example of internal structures construction. This file window.dbf is an ordinary .dbf structure and each row of it corresponds to a particular window description. A fragment of fields:

Field_name	Functionality
W_KEY	Key list
W_JOINT	Searching area
W_GET_FLD	Getting
W_GET_EXPR	from the connection
W_PUT_FLD	Putting
W_PUT_EXPR	into the connection
W_SUBMISS	Hierarchy of submission of files
W_SUB_FLD	Submission field
W_SUB_EXPR	expression
W_SUB_COND	condition

Figure 1: Structure of "Submission".

The first group controls the automatic searching facilities of documents entered in different workstations. Then we search by several Key expressions in the sets of indexed files. And in case we find the related document from the Search area we incorporate the information within this two document forms.

The second group makes a hierarchy of related files and defines the expressions and conditions of sharing information within the movement through this structures.

The temporary consistence of fields might be:

Field_name	Description
W_KEY	{{"NACC","DACC"}}
W_JOINT	{"Material registration"} {'{["sum_acc","VAT","freight",
	"reject","quan","sum"]}'}
W_GET_EXPR	{"{'quansum'}"}

Figure 2: Fragment of "Joint".

An analog mechanism is included for control of special dictionaries, containing standards and common information. Selection of data from this dictionaries simplifies the issues of providing integrity of data. Algorithmically these mechanisms are based on pattern recognition and the best matching searches. The fact is that the same information is provided differently in documents specially when it is an outside source. As a rule there are a lot of mistakes in these information. There might be also the inputting mistakes. So the problem is the selection of the best matches (on the bases of appropriate definitions and metrics) which is the starting point of the final selection of the corresponding term from the dictionary.

Therefore, the described system is designed and made out as an universal and flexible tool. It can be implemented in the particular subject area in comparably little time. It can be improved specific parts of the application without changing in the main modules, i.e. the system is independent not only from data format, but also from the subject area, as.

Resuming the work we can conclude that the Industrial Management Problem can be successfully solved on base of Generator Software Shells, which might be transformed into the local workstations completing the corresponding information structures and parameters externally. This is an alternative approach to the idea of integration into a global software solution diverse software packages working effectively as stand-alone programs. The additional algorithmic tools - the pattern recognition algorithms, selected on base of corresponding optimization and the flexible searching tools of exact and partial matching and others highly increase the intellectual property of application software systems of these kind.

References

- L. H. Aslanian. The discrete isoperimetric problem and related extremal problems for discrete spaces. Problemy Kibernetiki, Volume 36, pages 85-128, 1989.
- [2] L. H. Aslanian. The discrete modeling technologies. In I International Conference on Application of Critical Technologies for the Needs of Society, pages 160-161, Yerevan, September 1995.
- [3] L. H. Aslanian and Ja. Demetrovics. Closures, uniquely definably by a set of minimal keys. In II All-Union Conference: Informatics - 87, pages 160-161, Yerevan, 1987.

- [4] L. H. Aslanian and V. M. Karakhanian. The pattern recognition methods for information retrieval systems. In 7-th All Russia conference: Mathematical techniques and Optimization in Pattern Recognition, pages 25-30, Pushchino, September 1995.
- [5] S. Atre. Data Base: Structured Techniques for Design, Performance, and Management with Case Studies. John Wiley & Sons, 1980.
- [6] D. Maier. The Theory of Relational Databases. Computer Science Press, 1983.
- [7] V. Mařík and J. Lažanský. Applying AI for the development of CIM Software. In 13th World Computer Congress 94, pages 416-423, Elsevier Science B.V. (North-Holland), 1994.
- [8] R. L. Rivest. On the optimality of Elias's algorithm for performing best-match searches.
 In Information Processing 74, pages 678–681, Nort-Holland Publishing Company, 1974.
- M. Stonebraker, P. M. Aoki, R. Devine, W. Litwin and M. Olson. Mariposa: A new architecture for distributed data. In Sequoia 2000 Technical Report 93/31, University of California, Berkeley, 1993.
- [10] A. Tictäväinen and A. Perko. There are no unknown perfect binary codes. Anhales Universitatis Turkuensis, Volume 148, Number sor. A., 1971.
- [11] Yu. I. Zhuravlev. An algebraic approach to the problem of pattern recognition and classification. Problemi Kibernetiki, Volume 33, pages 5-68, 1978.
- [12] V. A. Zinoviev and V. K. Leontiev. On the perfect codes. Problems of Information Transfering, Volume 8, Number 1, pages 26-35, 1972.