

BULLETIN D'INFORMATIQUE APPROFONDIE ET APPLICATIONS

N° 36 DECEMBRE 1993

COMITE SCIENTIFIQUE

*France Chappaz
M'hamed Charifi
Roger Cusin
Bernard Goossens
Patrick Isoardi
Jean - Michel Knippel
Jean - Philippe Lehmann
Patrick Sanchez
Rolland Stutzmann*

DIRECTEUR

Edmond Bianco

REDACTEUR EN CHEF

Edmond Bianco

1 EDITORIAL,

par Edmond Bianco

REDACTEUR ADJOINT

Sami Hilala

**3 BIDIRECTIONAL, TWO DIMENSIONAL
REGULAR DATA FLOW ARCHITECTURE,**

par Jalal Almhana

13 ANALYSE AUTOMATIQUE DE REQUETES,

par Mohammed Taïeb Laskri

29 VOZZAVEDIBISAR,

par Jean - Michel Knippel

Publication gratuite trimestrielle de l'Université d'Aix - Marseille II
58, boulevard Charles Livon. F - 13284 Marseille Cedex 07
Téléphone : (33) 91 39 65 00 Télécopie : (33) 91 31 31 36

Edition 1995

ISSN 0291 - 5413

ÉDITORIAL,

Réalité et irréalité.

Dans les années cinquante à Marseille, une puanteur insupportable recouvrait la ville, lorsque par temps mou les fumées stagnaient. Et puis un jour une très violente explosion souffla le quartier de Rabatau. L'extraction de l'huilerie qui polluait la cité venait de sauter. Les cadavres du personnel de l'usine retrouvés cuits à point, de nombreux appartements dévastés aux alentours, des quantités de blessés, toutes les vitres à la ronde éclatées, bref de quoi causer une intense émotion. Or il se trouve que l'émotion est une denrée riche, en ce sens qu'elle est d'une exploitation remarquable. La presse et divers organismes dégoulinèrent alors de bons sentiments en pleurant sur les malheurs qui s'étaient abattus brutalement de la sorte sur une masse de pauvres gens. Et au nom de ces pauvres gens on organisa une collecte. Qui permit de réunir une somme rondelette.

Et puis après me direz-vous ? Après rien. L'argent se perdit dans on ne sait trop quels sables administratifs ou privés. En tout cas les " pauvres gens victimes de la catastrophe " et au nom de qui la collecte avait été faite n'en virent jamais un centime. Y eut-il peut-être un scandale, qu'il fut vite étouffé.

Il y eut également, toujours dans les années cinquante, le drame de Fréjus. Un barrage qui lâche et une énorme vague qui ravage toute une vallée. Là encore de nombreuses victimes. Là encore un troupeau lamentable de pauvres gens ayant tout perdu quand ils n'avaient pas en prime perdu la vie. Émotion. Collecte. On réunit une somme fantastique pour l'époque, de l'ordre de dix milliards, et toujours et encore ce furent les plus nécessiteux qui ne virent jamais la couleur de cet argent, qui, encore une fois se perdit dans les sables insatiables de l'administration ou des organismes aussi puissants qu'au dessus des lois. Devant l'importance du scandale il y eut de nombreux procès ... encore inachevés, quarante ans plus tard. On comprend d'ailleurs parfaitement que la Justice s'aiguise les dents sur les voleurs d'oranges et de pots de yaourts justiciables qu'il ne faut ni faire lanterner ni décevoir, et prenne largement son temps pour aiguïser ses couteaux à l'odeur d'un racket de cette ampleur. Pensez, les enquêtes sont forcément longues et délicates et il y a des pieds sur lesquels il convient de ne pas inconsidérément marcher.

Je revois encore les efforts de mes instituteurs, cela remonte aux années trente et quarante, pour essayer de nous inculquer cette honnêteté du citoyen, base de toute éducation républicaine et laïque. Leurs efforts que j'ai de plus en plus tendance à considérer comme méritoires n'auraient-ils pas abouti à faire de nous des sortes de rêveurs qui vivent dans l'irréel, crédules face aux déclarations péremptoires autant que savamment médiatisées, qui votent pour des " Mon général " et jusqu'à des " Savate Premier ", lesquels nous promettent la lune avant et nous offrent des " bombinettes ", après, à nos frais bien sur et dont le financement ne semble jamais offrir les mêmes difficultés que celui de la Sécurité Sociale. Allez savoir pourquoi.

De mauvais esprits vont même jusqu'à prétendre que tous les vrais problèmes, ceux qui concernent la vie et la santé des citoyens sont systématiquement rejetés vers la charité publique, qu'on aiguillonne en entretenant une forte inquiétude médiatique. Et rajoutent-ils en un dernier trait encore plus vénéneux, cela donne l'occasion d'un deuxième pillage, les fonds récoltés n'allant jamais à leur destination. Vous vous rendez compte ? Et on rajoute que les Comptables officiels, si sourcilleux quand il s'agit des "biens sociaux" ou des "biens de l'État", le seraient bizarrement beaucoup moins quand il s'agit de détournement de collecte "humanitaire". On objectera que ces "histoires" tout de même datent un peu. Les années cinquante c'est loin. On parie?

Un petit bonhomme tout rond, à la voix émue, au ton ardent, et même frémissant, familier du petit fénestron, vous invite à "participer" à la recherche contre le cancer, en finançant son Organe. Quoi de plus émouvant à part peut être la lutte contre le Sida ? Et puis on s'aperçoit que seule la portion congrue parvient, et encore peut-être, à sa destination déclarée. Et l'on peut voir de dignes Professeurs chenus et respectables faisant partie du brain-trust de l'Organe, venir déclarer penauds qu'ils n'ont rien vu passer.

Quand je suis à table je n'aime pas que mon chien assiste au repas car, bavant et l'œil brillant il ne peut s'empêcher de mendier quelques miettes. Néanmoins il y a des exceptions, et il m'arrive de faiblir jusqu'à lui jeter quelque peau de saucisson ou quelque os de poulet, moyennant quoi il se tait.

Mais au fait, pourquoi au milieu d'un sujet aussi sérieux me mets-je à vous parler de mon Chien ?

Même le juteux Téléthon semble quelque peu mité, on aurait déjà vu un trou. Aussi de ma modeste voix je conjure les forces judiciaires de se boucher hermétiquement les yeux, "des fois" qu'on s'apercevrait que tout l'argent ne va pas à l'enfant malade ...vous vous rendez compte d'une désillusion. Non, non, il ne faut pas aller y voir ! Ce saut de l'irréalité dans la réalité serait bien trop horrible.

Science et Conscience, vieilles notions surannées. Toute la logique du monde libéral repose sur la compétition, le profit, axiomes de base. Conséquence première: qu'importe ce que l'on fait, pourvu que cela se vende bien. Et quoi de meilleur en ce domaine que le vent, qui n'exige entre autres aucun SAV, (service après vente).

Heureusement Internet, qui se vend bien aussi, nous promet une cyberréalité irréelle autrement plus gratifiante que cette sordide réalité.

E. Bianco

BIDIRECTIONAL, TWO DIMENSIONAL^o REGULAR DATA FLOW ARCHITECTURE

Jalal Almhana
Computer Sciences Department,
University of Moncton, Moncton,
NB, Canada E1A 3E9

Phone No. (506) 858-4335

e.mail: ALMHANJ@UMONCTON.CA

Abstract

We describe a new, static, cellular data flow architecture with emphasis on its function mode and exchange means between cells; an algorithm is defined for instruction packet routing. We also define the neighbouring concept and implementation rules which permit optimum mapping.

I. Introduction

In the last few years, Data Flow Machines (DFM) have had remarkable success due to the high level of performance that can be obtained in comparison with other more conventional machines. Many DFM have been studied, some of them have been simulated [6], [14], [20], [1], some others have been implemented [7], [16], and a number of prototypes have been proposed [3], [11], [17], [9]. The DFM are based on the following principles [5], [10], [8], [22], [13]; 1) A data flow operation is enabled if, and only if all the required operands become available. This means that computation is data driven control in asynchronous manner. 2) A data flow operation is purely functional and produces no side effects. Most simulated and realized DFM [11], [17], [4], [15], whether static or dynamic, do not have a regular structure or a system of communication simple enough to facilitate their realization in VLSI technology. These characteristics have made a success of the systolic machines [18], [19], which also use a similar concept: Data synchronization and spatial parallelism as defined by an algorithm graph. Our objective is to define a static data flow architecture that has a simple structure and an efficient means of communications. This would be possible through VLSI Technology without limiting it to a certain number of algorithms, as is the case with systolic machines. The following points will be studied in this paper: Description of the architecture proposed; algorithm implementation in the architecture; Information exchange between cells and execution of a program; and finally, external communication exchange and architecture extension.

II. Description of the architecture proposed

Fig. 1 shows the architecture proposed. An architecture of 5x5 cells is used as an example. This structure is based on an assembly of cells arranged in a square shape. Intercellular communications are bidirectional in two dimensions. Each cell is composed of two parts (Fig. 2-a).

- 1- The upper part, called the communications unit.
- 2- The lower part, called the processor.

© Copyright, Jalal Almhana, May 1991.

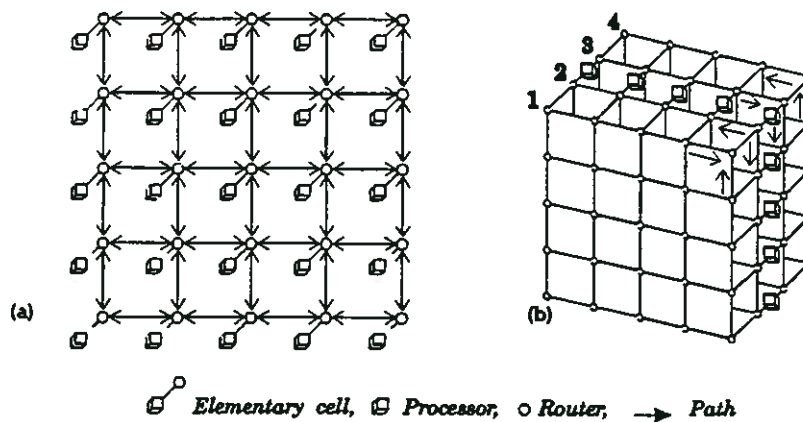


Fig.1 a- A bidirectional architecture with two dimensions.
 b- 1,2,3,4 unidirectional routing networks.

The communications unit is composed of four identical unidirectional routers which belong to four independent networks: x^+y^+ , x^-y^- , x^+y^- , x^-y^+ . (Fig 6) These together could be seen as one bidirectional network in two dimensions. We intentionally chose a unidirectional router for many reasons: 1- To simplify its structure and then facilitate its integration in VLSI; 2- To be able to find a simple routing algorithm; 3- To improve the pass band of the communication network. Each router is capable of receiving data from 3 inputs (x_i, y_i, z_i) and routing data through three pipelines to 3 outputs (x_o, y_o, z_o), see Fig. 2-b. z_o, z_i , represent the input and the output of the processor's router. x_i, x_o, y_i, y_o represent communication inputs and outputs with the other neighbour routers on x, y coordinates.

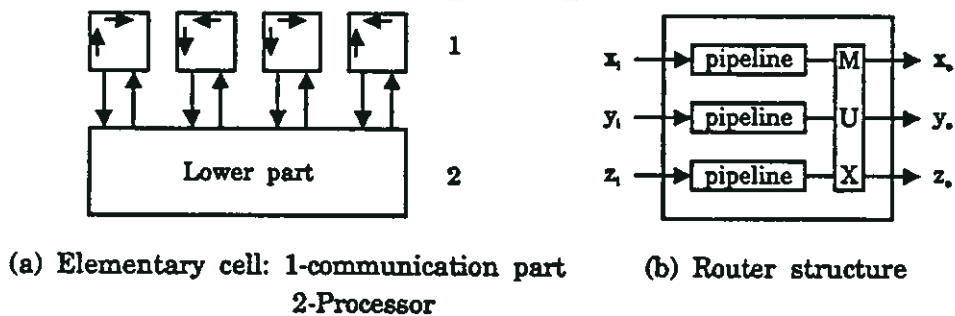


Fig2.

The lower part of the cell, Fig. 2-a, may represent in the most general case: 1- a non-specialized processor with a local memory; 2- a specialized processor (functional unit) with a local memory; 3- a an input/output interface for external communications. Each processor is capable of executing an instruction packet and performing a result packet. Recall that an instruction packet includes an operation code, operande values, and destination fields that specify where results should be sent. A result packet consists of a data value and a destination field. In our case the destination address is divided into two parts: the cell address and the local memory address.

Like all other data flow machines, our architecture does not have a central memory but a locally distributed memory. Its size is, in general, small in order to allow high speed access. It has two ports A & B and each may be used as a read port, while one can also be used as write port. Data from A and B can be read simultaneously by specifying the appropriate addresses on the A and B address lines (Fig. 3). The packet Read/Write circuit is responsible for packet exchange between the router or processor and the local memory.

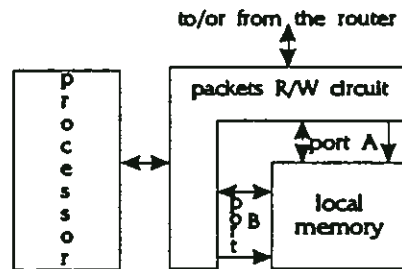


Fig. 3 Lower part structure

III. Algorithm implementation in the architecture

This operation is achieved by the compiler which distributes the algorithm tasks throughout the cells and whose number is already known. The complexity (granularity) of the task is variable and has yet to be defined. Though much research has already been done in this area, the problem is still unresolved. The mapping operation must :

- Assign each task or set of operations to a cell. In other words, to determine the address of the cell and its local memory address.

- Specify the destination addresses for partial results. These are determined by the interdependence imposed by the calculated algorithm.

To minimize the data path length between two cells, it is best to respect a concept that we will call "neighbouring". This is very important in reducing the communication time between cells. It permits local and distributed transfers through the communication networks, which help to avoid bottleneck problems and improves the global performance of the communication system. Unfortunately, this principle is often ignored in research concerning data flow program mapping. The architecture proposed in this paper is designed to favour "neighbouring". Our intention, in this paper, is not to delve too deeply into the intricacies of the problems of mapping, but to establish some allocation rules that we consider useful in improving transfer time.

-Neighbouring concept

The neighbouring concept includes three definitions:

Definition 1: Spatial neighbouring

Two tasks are said to be Spatially Neighboured (SN) if, and only if, they are linked by an arc on the associated algorithm graph.

In other words spatial neighbouring means that the result of one task is used as data for the other task.

Definition 2: Temporal neighbouring

Two tasks are said to be Temporally Neighboured (TN) if, and only if, the beginning of one task coincides with the end of the other tasks.

Definition 3: Spatio-temporal neighbouring

Two tasks are said to be Spatio-Temporally Neighboured (STN) if, and only if, they are both SN and TN.

To illustrate these definitions, let's consider the following example:

Example 1

Let's calculate the formula: $(c+d)/(c*d)/a*b$. Its associated graph is shown in figure 4.

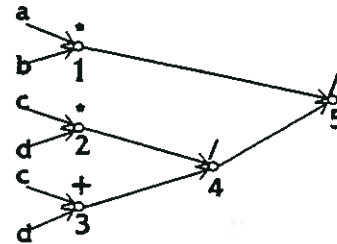


Fig.4 Dependencies graph

Suppose that multiplication time is $5 \mu s$, addition time is $3 \mu s$, and division time is $7 \mu s$. We can deduce the following table:

| Task | Predecessor | Successor | Beginning time | Ending time |
|------------------|-------------|-----------|----------------|-------------|
| 1: $o_1=a*b$ | - | 5 | 0 | 5 |
| 2: $o_2=c+d$ | - | 4 | 0 | 3 |
| 3: $o_3=c*d$ | - | 4 | 0 | 5 |
| 4: $o_4=o_2/o_3$ | 2,3 | 5 | 5 | 12 |
| 5: $o_5=o_4/o_1$ | 1,4 | - | 12 | 19 |

Table 1.

Using the definitions given earlier (1,2,3), we can identify three groups of neighbour tasks:

$G1 = \{ (1,5), (2,4), (3,4), (4,5) \}$: Tasks SN

$G2 = \{ (1,4), (3,4), (4,5) \}$: Tasks TN

$G3 = \{ (4,5), (3,4) \}$:Tasks STN .

The spatial neighbouring may be determined directly from the associated algorithm graph. On the other hand, temporal neighbouring may only be determined by using simulation. To accomplish this, task execution time should already be known. This kind of simulation is very well known and is usually used to define the critical path and margin of time relative to execution of tasks.

Definition 4: Optimum mapping

Associated algorithm graph mapping is said to be optimum if, and only if, the data transfer time between the cells involved in a particular task is equal to, or less than the margin of time relative to the critical path.

According to this definition and those relative to concept of neighbouring, optimum mapping implies the following assignation rules:

1- Two tasks STN should be assigned to two neighbour cells (or to the same cell if this does not affect parallelism). We assume that the transfer time between neighbour cells is so slight that it can be considered "equal" to the minimum margin time (zero) or included in the task time.

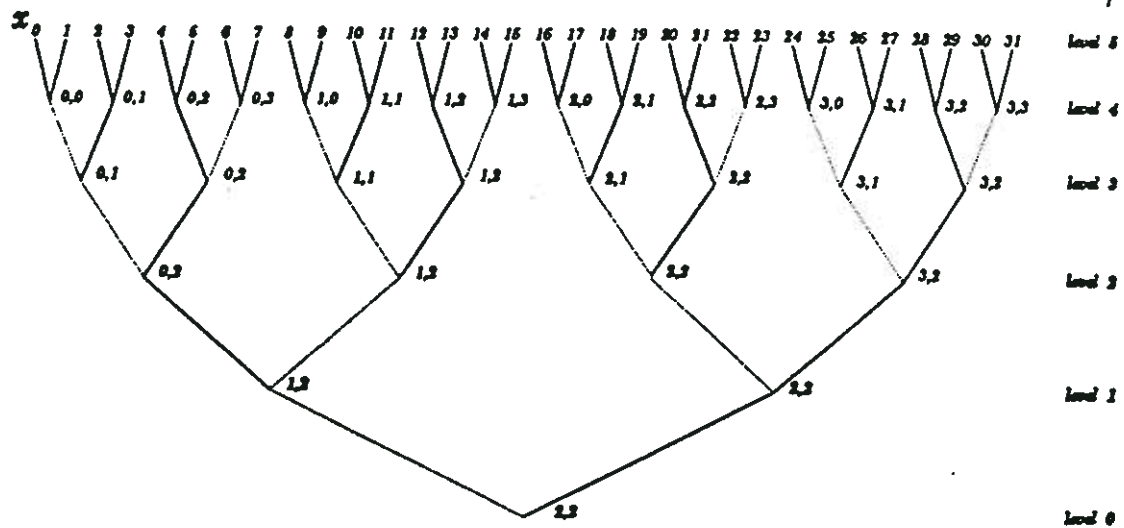


Fig. 5-a Binary tree of the sum

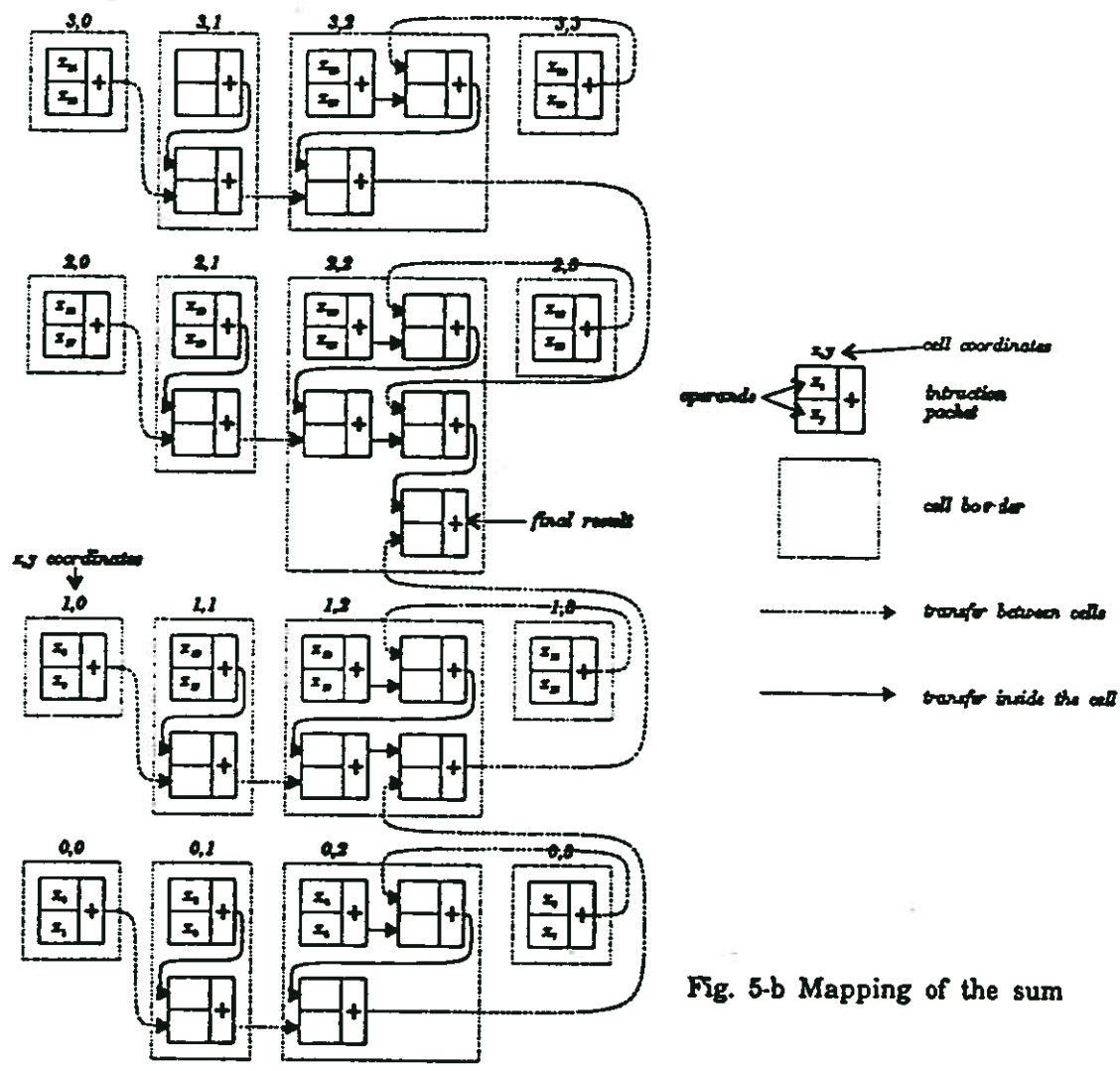


Fig. 5-b Mapping of the sum

2- Two tasks SN should be assigned to two cells where the transfer time between them is equal to or less than the margin time.

Let's note that systolic machines respect these rules because their structure is deduced from the associated algorithm graph, but in many cases, [18][19], an added delay is necessary between tasks to achieve spatio-temporal synchronization. In the architecture proposed in this paper, these added delays are not necessary, as data flow machines are asynchronous and operations are synchronized on the presence of data. In this, paper we will limit ourselves to studying an example of optimum mapping using the above rules. In future works, we will study the possibilities and the methods that will lead to optimal mapping.

Example 2

Let's calculate the sum S of $N=32$ values a_i , $i=1.....32$. The figure(5-a) shows the graph which represents the computation of the sum using a binary tree. The number beside each node (task) shows the x,y coordinates of the cell where this task is mapped. Dotted lines indicate where there is a result transfer between cells, while continuous lines indicate that the results will be kept in the same cell for use of successor operations. All tasks are similar (addition operations) and have the same execution time. This fact implies that all SN tasks are also TN and should be assigned to neighbour cells according to the assignation rules. This graph may be represented by a group of packets mapped into the architecture. Figure(5-b) shows the result of this mapping and the relations between the packets. Notice that all data transfers are confined to the cell or neighbour cells involved: no linking packet arcs between cells that are not neighbours. Such mapping provides maximum parallelism with minimum data transfer time.

IV. Information exchange between cells and execution.

First, we assume that all cells have specific addresses determined by their x, y coordinates. Each cell is capable of receiving or transmitting data (packets) to and from any other cell in the architecture. These operations are achieved by means of the four routing networks mentioned above, (Fig. 1-b). The sending of the packet involves two steps:

1- Selection of the routing network that allows the packet to reach the target cell, which depends on the source cell address (x_s, y_s) and the target cell address (x_t, y_t) . Fig. 6 shows an example of which target cells (bold) could be reached from a source cell, such as (1,3), according to the different networks.

2- Routing the packet through the selected network.

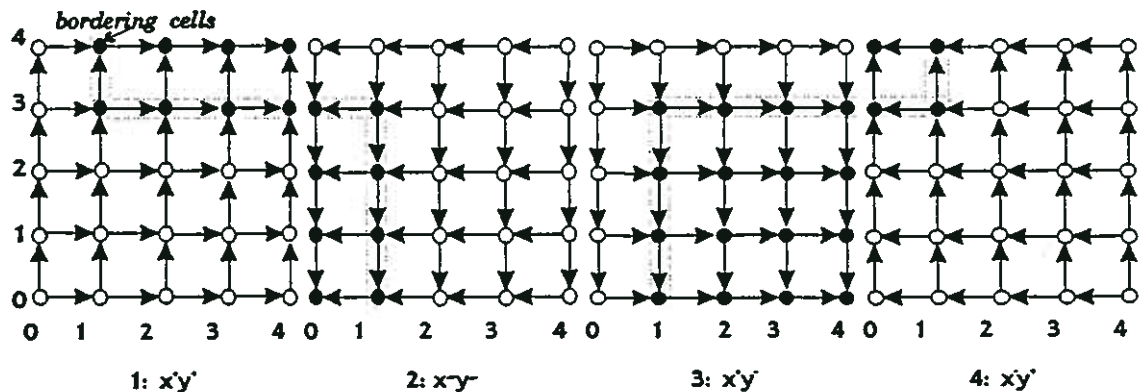


Fig.6 Four unidirectional networks are possible with two dimensions.

To achieve the first step, we will establish the following truth table(2) which is the result of comparisons between the source address (x_s, y_s) with the target cell address (x_t, y_t):

| a | b | c | d | a'b' | c'd' | Selected network | | | |
|---|---|---|---|------|------|------------------|----------------|----------------|----------------|
| | | | | | | n ₁ | n ₂ | n ₃ | n ₄ |
| 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 |
| 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 |
| 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 |
| 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 |
| 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |

Table 2

From this table, and after simplification, we can write the following boolean equations to determine which network should be used to reach the target cell :

$$\begin{aligned}
 n_1 &= ac + b'd + a'b'cd && \text{Select network 1: } x^+y^+ \\
 n_2 &= bd' + a'c && \text{Select network 2: } x^-y^- \\
 n_3 &= ab'd' + b'c && \text{Select network 3: } x^-y^+ \\
 n_4 &= a'd + a'b'c && \text{Select network 4: } x^+y^- \\
 \text{Whereas : } & a = y_t > y_s, \quad b = y_t < y_s, \quad c = x_t < x_s, \quad d = x_t > x_s, \quad a' = \text{Not } a
 \end{aligned}$$

Some cells, such as the bordering cells, could equally be reached through two networks. In this case, one network is selected arbitrarily.

The second step, (routing through a network), is achieved by comparisons of the target cell address with each cell address that is passed. We can write the following truth table (3):

| e | f | Routing | | |
|---|---|-------------------|-------------------|-------------------|
| | | to z ₀ | to x ₀ | to y ₀ |
| 1 | 1 | 0 | 1 | 1 |
| 1 | 0 | 0 | 1 | 0 |
| 0 | 1 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 | 0 |

Table (3)

From this table and after simplification, we can write the three boolean equations used to direct a packet towards the outputs x_0, y_0, z_0 .

$$\begin{aligned}
 x_0 &= e && \text{towards } x_0 \\
 y_0 &= f && \text{towards } y_0 \\
 z_0 &= e'f && \text{towards } z_0 \\
 \text{whereas: } & e = x_s < x_t, \quad f = y_s < y_t
 \end{aligned}$$

When the addresses match, the packet is sent to the processor, (z_0 output). In the cases where the packet could be sent equally towards x_0 or y_0 , the most available output will then be chosen. In this way,

the packets are directed from one cell to another, until they reach their destination cell. Once the packet reaches its target cell, it is stored according to its local memory address.

It could be demonstrated that the length of the paths taken by each packet is optimal and is equal to: $|x_s - x_t| + |y_s - y_t|$.

Execution

The enabled operations execution begins simultaneously in all the cells in an asynchronized manner. Each processor executes the enabled operations in its local memory and according to the destination address keeps the result in its local memory or sends it to another cell through the communication unit. The enabling rules are the same as those for data flow machines.

Execution is completed when there are no more enabled operations.

V. Input/Output (External Communications Exchange)

As stated above, the lower part of the cell could function simply as an input/output interface for external communication. The location of these Input/Output cells may be defined by the user according to his application needs. One could imagine, for example, a column of cells used as a window for external information exchange.

Such a window also makes enlargement possible by connecting many such architectures together, Fig.7.

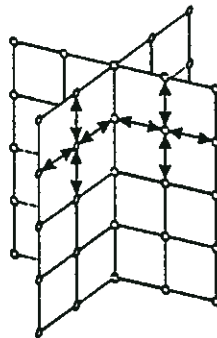


Fig.7 A three dimensional bidirectional architecture can be obtained by combining two of the two dimensional architectures.

VI. Conclusion

In this paper we described a new static DFA which has the following characteristics:

- a regular cellular structure and simple asynchronized communications, which would facilitate its realization in VLSI technology.

- Evolutive, permitting a complex machine based on simple structure cells.

- It is not limited to a certain number of algorithms, as is the case with systolic machines.

These characteristics should make high levels of performance possible.

A simple and efficient switching algorithm was defined. It produces an optimal path length for packets routing and it is easy to implement in hardware. We also defined the concept of neighbouring which minimizes data transfer time between cells.

In future research, we will explore in more detail the different possible methods leading to optimal mapping. we will also evaluate architecture performance by studying simulations based on the Data Flow Petri Nets (DFPN) model [2], that we defined in previous publication.

References

- [1] J. Almhana, "Réalisation d'un opérateur de traitement numérique rapide par une machine à flux de données," *Annales des Télécom.*, tome 39, no 7-8, pp.323-332, Jul. 1984.
- [2] -----, " Modélisation par réseaux de Petri à flux de données; application à la synthèse de l'opérateur de riccati rapide, " *Thèse d'état, Univ. Aix-Marseille III*, Juin 1983.
- [3] Arvind and V. Kathail, "A multiple processor data flow machine that supports generalized procedure," in *Proc. 8th Annu. Symp. Comput. Architecture*, pp. 291-302, May 1981.
- [4] Arvind and D. E. Culler, "Data flow architecture," *MIT/LCS/TM-294*, lab. For Comput. Sci., MIT, Feb. 1986.
- [5] J. Backus, "Can programming be liberated from Von Neumann style? A functional style and its algebra of programs," *Commun. ACM*, vol. 21, pp. 613-641, Aug.1978.
- [6] W.W. Carlson and K.Howang, "Algorithmic performance of data flow multiprocessor, " *IEEE Comput. Mag.*, vol. 18, no. 12, pp. 30-40, Dec.1985.
- [7] D. Comte, N. Hifid, and J. C. Syre, "The data driven LAU multiprocessor system: Results and perspectives, " in *Proc. IFIP Congress, 1980*, pp. 175-180.
- [8] A. L. Davis and R. M. Keller, "Dataflow program graphs, " *IEEE Comput. Mag.*, vol. 15, no. 2, pp. 26-41, Feb. 1982.
- [9] J.B. Dennis and G.R. Gao, "An efficient pipelined dataflow processor architecture, " *Joint Conf. on Supercomputing*, pp. 368-373, *IEEE Comput. Society and ACM SIGARCH*, Florida, Nov. 1988.
- [10] J.B. Dennis, "Dataflow supercomputers, " *IEEE Comput. Mag.*, vol. 13, no. 11, pp. 362-376, Nov. 1980.
- [11] J.B. Dennis and D. P. Misunas, " Apreliminary architecture for a basic dataflow processor, " in *Proc. 2nd Annu. Symp. Comput. Architecture*, pp. 126-132, Jan. 1975.
- [12] J. L. Gaudiot, Y. H. Wei, "Token relabeling in a tagged token dataflow architecture," *IEEE Trans. Comput.*, vol.38, no.9, Sept. 1989.
- [13] D. Ghosal, L. N. Bhuyan, " Performance evaluationof a dataflow architecture, " *IEEE Trans. Comput.*, vol. C-39, no. 5, pp. 615-627, May 1990
- [14] K.P. Gostelow and R. E. Thomas, "Performance of simulated dataflow computer," *IEEE Trans. Comput.*, vol. C-29, no. 10, pp. 905-919, Oct. 1980.
- [15] J. R. Gurd and I. Watson, "A practical dataflow computer, " *IEEE Comput. Mag.*, vol. 15, no. 2, pp. 51- 58, Feb. 1982.
- [16] J.R. Gurd, I. Watson and C.C. Kirham, "The Manchester prototype dataflow computer, " *Commun. ACM*, vol. 28, pp. 34-52, Jan. 1985.
- [17] R. M. Keller, G. Lindstrom, and S. Patil, "A loosely coupled applicative multiprocessing system, " in *IFIPS Conf. Proc.*, Vol. 48, pp.613-622, June 1979.
- [18] H.T. Kung, "Why systolic architecture, " *IEEE Comput.*, vol. 15, no.1, pp. 37-46, 1982.
- [19] F.C. Lin and K. Chen, "On the design a unidirectional systolic array for key enumeration, " *IEEE Trans. Comput.*, vol. C-39, no2, pp. 266-269, Feb.1990.
- [20] L. M. Patnaik, R. Gvindarajan and N. S. Ramdoss, "Design and performance evaluation of EXMAN: An extended Manchester datafolw computer," *IEEE Trans. Comput.*, vol C-35, no. 3, pp. 229-243, Mars. 1986.
- [21] P. Rong Chang and C.S. G. Lee, "A decomposition approach for balancing large-scale acyclic data flow graph, " *IEEE Trans. Comput.*, vol. 39, no. 1, pp.34-46, Jan. 1990.
- [22] P. C. Treleave, B. R .David, and P.R. Hopkins, "Data driven and demand driven computer architecture," *Comput. Surveys*, vol. 14, pp.93-143, Mar. 1982.

ANALYSE AUTOMATIQUE DES REQUETES PASSAGE DU LANGAGE NATUREL AU LANGAGE DOCUMENTAIRE

M. T. LASKRI
Université de Annaba
Institut d'Informatique
G.R.I.A. - B.P. 12
DZ - 23200 ANNABA

Résumé

Le langage n'a cessé au cours du temps de soulever l'interrogation. Le développement des langages formels, les travaux des logiciens, depuis Turing, ont montré la voie à l'automatisation du discours.

En même temps les linguistes révélaient par leurs recherches les structures immanentes aux langues naturelles.

L'informatique dote une machine de tous les appareils du langage. Il s'agit de permettre à la machine d'atteindre au langage des hommes et d'en automatiser les traitements les plus divers:

- traduction du langage naturel en langage documentaire,
- structure des requêtes,
- compréhension des requêtes,
- relations entre descripteurs,
- distance entre descripteurs.

Ce séminaire présente l'analyse des systèmes documentaires et la démarche choisie pour ces automatisations.

J. - M. KNIPPEL

ANALYSE DES SYSTEMES DOCUMENTAIRES

1°) Inconvénients de la Classification Décimale Universelle (CDU)

- * EXPLOSION DE L'INFORMATION DOCUMENTAIRE
- * LES DOCUMENTALISTES NE PEUVENT PAS ETRE SPECIALISTES DANS TOUS LES DOMAINES
- * LA QUANTITE TEND A NOYER LA QUALITE DES SERVICES D'UN DOCUMENTALISTE

L'INFORMATIQUE PERMET DE MULTIPLIER LES CHEMINS D'ACCES A UN DOCUMENT

2°) Automatisation de la gestion du prêt

Problème de gestion relativement simple :

- * Réservations
- * Edition des lettres de réclamation
- * Effectuer des analyses statistiques

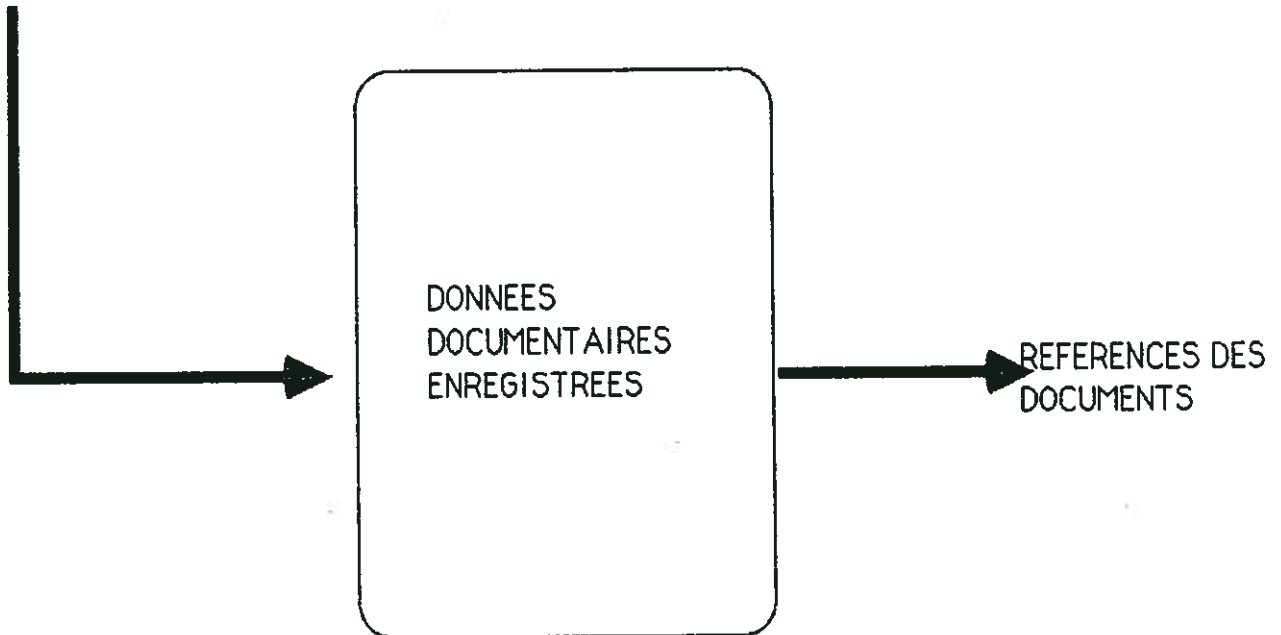
3°) Automatisation des recherches documentaires

ROLE PRINCIPAL D'UN SYSTEME DOCUMENTAIRE

SIMILAIRE

A CELUI DES DOCUMENTALISTES

REQUETES DES
UTILISATEURS



LES TRAVAUX S'ACCENTUENT SUR :

- * COMPREHENSION DES REQUETES
- * REPRESENTATION DU CONTENU DES DOCUMENTS DU FONDS DOCUMENTAIRE
- * DEFINITION D'UNE FONCTION DE RECHERCHE

LE SYSTEME DOCUMENTAIRE NE DOIT PAS SEULEMENT ATTEINDRE LES DOCUMENTS INDEXES PAR LES MEMES TERMES QUE CEUX DE LA REQUETE

MAIS

EGALEMENT AVEC DES TERMES AYANT UN CERTAIN RAPPORT ETROIT AVEC LA REQUETE

EXEMPLE :

SOIT LA REQUETE :

" CULTURE, BLE, ALGERIE"

SOIT UN DOCUMENT INDEXE PAR :

"CULTURE, CEREALES, MAGHREB"

POUR ATTEINDRE CE DOCUMENT, LA REQUETE DOIT ETRE ELARGIE POUR DEVENIR :

| | | |
|----------|----------|---------|
| CULTURE, | BLE | ALGERIE |
| | OU | , OU |
| | CEREALES | MAGHREB |

PRENDRE EN CONSIDERATION LES DIFFERENTES RELATIONS ENTRE TERMES

SYNONYMIE, HIERARCHIE, ASSOCIATION



L'ENSEMBLE DE CES RELATIONS CONSTITUE UN THESAURUS

DEFINITION D'UN THESAURUS SELON LES SYSTEMES :

Pour SATIN, SYNTOL :

UN THESAURUS EST UN DICTIONNAIRE QUI DEFINIT LES CORRESPONDANCES ENTRE LES MOTS CLES DU LANGAGE DOCUMENTAIRE ET LE VOCABULAIRE DU LANGAGE NATUREL

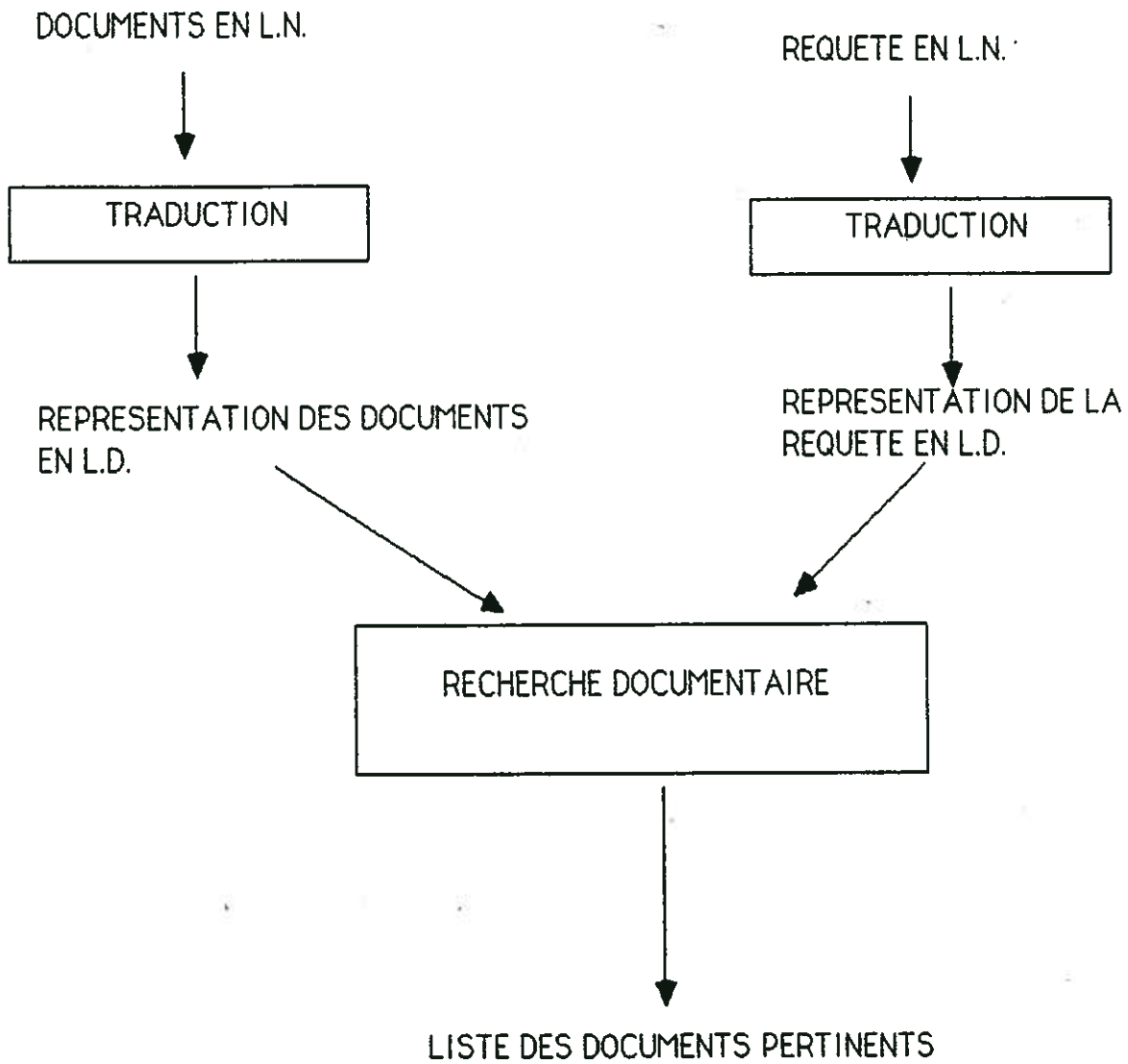
Pour MISTRAL, SMART :

UN THESAURUS REPRESENTE UN DICTIONNAIRE QUI DEFINIT DES LIENS SEMANTIQUES ENTRE LES DESCRIPTEURS DU LANGAGE DOCUMENTAIRE

Pour SYSUT :

UN THESAURUS EST UNE FORME ORGANISEE FACILEMENT ACCESSIBLE CARACTERISEE PAR DES LIENS SEMANTIQUES ENTRE DESCRIPTEURS.

TRADUCTION L.N. → L.D.



**PRINCIPAUX PROBLEMES RENCONTRES LORS DE LA TRADUCTION
AUTOMATIQUE DES REQUETES ET DES DOCUMENTS EN LANGAGE
DOCUMENTAIRE**

**a) reconnaître les différentes formes graphiques
(Analyse morphologique)**

b) reconnaître les groupes de mots indissociables

EX : Moelle épinière

Maison d'arrêt

Chemin de fer

...

c) Interpréter les termes polysémiques :

**la traduction d'un terme du L.N. dans le L.D. varie en
fonction du contexte du terme**

EX: Groupe Sanguin (en Médecine)

Groupe Abélien (en Mathématiques)

...

**d) Pouvoir confronter le texte à des connaissances qui sont
présentées suivant un modèle pré-défini dans le domaine de la
compréhension d'énoncés écrits**

*** Déterminer la structure profonde d'une phrase
(les modèles syntaxiques)**

*** Shank considère qu'un texte peut être représenté par
un graphe (théorie de la dépendance conceptuelle)**

*** Coulon et Kayser : comprendre un texte revient à trouver pourquoi il
est plausible dans le contexte où il se trouve**

DEMARCHE POUR ANALYSER LES REQUETES

1. FILTRAGE DES DESCRIPTEURS D'UNE REQUETE

UN BAGAGE CULTUREL LINGUISTIQUE EST NECESSAIRE AU SYSTEME POUR COMPRENDRE LES REQUETES ET LE FONDS DOCUMENTAIRE

POUR SELECTIONNER LES DESCRIPTEURS D'UNE REQUETE : LE SYSTEME LIT MOT A MOT LA REQUETE : LES TERMES APPARTENANT A L'ENSEMBLE DES DESCRIPTEURS POUVANT SERVIR A L'INDEXATION, SONT INCLUS DANS LA LISTE DES DESCRIPTEURS DE LA REQUETE.

EX: JE VOUDRAIS UN LIVRE D'EXERCICES DE PROLOG ?

COMMENT ORGANISER LES DONNEES D'UN FICHIER ?

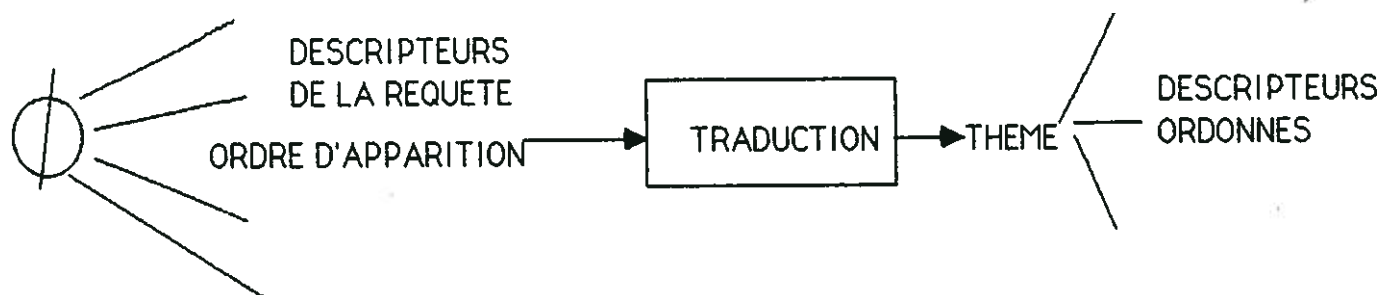
2. IMPORTANCE DES DESCRIPTEURS D'UNE REQUETE

LES DESCRIPTEURS D'UNE REQUETE SONT PLUS OU MOINS IMPORTANTS ET LEUR IMPORTANCE EST TRES DIFFICILE A DETERMINER.

UNE PONDERATION MANUELLE EST A EVITER CAR LA VALEUR D'UN POIDS A ATTRIBUER A UN DESCRIPTEUR D'UNE REQUETE VARIE.

ORDRE DE PRIORITE ENTRE LES DESCRIPTEURS D'UNE REQUETE A L'AIDE DE SON ANALYSE SYNTAXIQUE

LA REQUETE EST TRADUITE EN UN GRAPHE AYANT POUR RACINE LE THEME



RACINE VIDE SIGNIFIE QUE LE THEME EST INITIALEMENT INDETERMINE

REGLES D'ANALYSE SYNTAXIQUE

DETERMINATION DE LA STRUCTURE D'UNE REQUETE D'APRES L'APPARTENANCE DES DESCRIPTEURS A CERTAINES CATEGORIES SYNTAXIQUES OU A CERTAINES CLASSES SEMANTIQUES.

A) PREPOSITION en

traduction de : DESCi en DESCj

1) si DESCj est une date ---> DESCi ——— DESCj / date

EX: les ordinateurs en 1990

2) si DESCj est un lieu ---> DESCi ——— DESCj / lieu

EX: l'informatique en Algérie

3) si DESCi est une activité et si DESCj est une langue ou un langage --->

DESCi ——— DESCj / manière

EX: traduction en arabe
programmation en pascal

4) si DESCi est un déplacement et si DESCj est un moyen de transport --->

DESCi ——— DESCj / moyen

EX: un voyage en train

5) si DESCi est un travail et si DESCj est une durée --->

DESCi ——— DESCj / durée

EX: compilation en 25 secondes

(...)

B) PREPOSITION sur

DESC_i sur DESC_j

1) si DESC_i est un traitement informatique et si DESC_j est un outil informatique --->

DESC_i ——— DESC_j / moyen

EX: la comptabilité sur un ordinateur

(...)

C) PREPOSITION de

DESC_i de DESC_j

1) si DESC_i est une activité ou une action ---> DESC_j ——— DESC_i

EX: L'analyse d'un texte

(...)

D) PREPOSITION dans

DESC_i dans DESC_j

1) si DESC_j est un lieu ou un objet ---> DESC_i ——— DESC_j / lieu

EX: l'informatique dans les entreprises

(...)

E) PREPOSITION par

DESC_i par DESC_j

1) si DESC_i est traitement et si DESC_j est un outil --->

DESC_i ——— DESC_j / moyen

EX: traduction par le compilateur

(...)

EXEMPLES D'ANALYSE SYNTAXIQUES

1ère REQUETE : Description des ordinateurs de gestion

a) ordinateur est le thème --->

ordinateur ——— description

b) gestion devient le thème --->

gestion ——— ordinateur ——— description

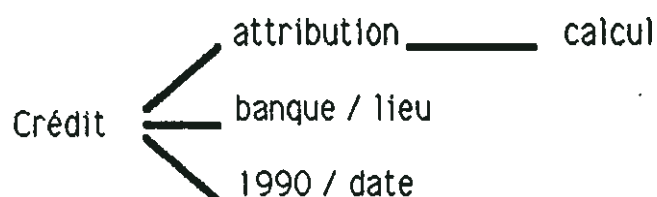
2ème REQUETE : Description de l'informatisation du prêt dans les bibliothèques

a) prêt ——— informatisation ——— description

b)



3ème REQUETE : Calcul d'attribution de crédits dans une banque en 1990 ?



NECESSITE DE L'ANALYSE SYNTAXIQUE

SOIENT "METHODE", "ANALYSE", "SYSTEME", "EXPLOITATION" QUATRE DESCRIPTEURS APPARTENANT AU LEXIQUE

A) LES DEUX TERMES "ANALYSE" ET "METHODE" NON ORDONNES PEUVENT AVOIR DEUX SENS :

- * ANALYSE D'UNE METHODE
- * METHODE D'ANALYSE

B) POUR "EXPLOITATION" ET "SYSTEME" :

- * EXPLOITATION DU SYSTEME
- * SYSTEME D'EXPLOITATION



L'ODRE DES TERMES PERMET DE CHOISIR LA SIGNIFICATION REELLE

LIMITATIONS DE L'ANALYSE SYNTAXIQUE

A) LES REQUETES EN L.N. PEUVENT S'EXPRIMER DE MULTIPLES FAÇONS
DONC IL EST DIFFICILE DE CONSTRUIRE UN ANALYSEUR SYNTAXIQUE POUR
L'ENSEMBLE DES REQUETES.

B) LES DONNÉES SEMANTIQUES ENTRE DEUX TERMES PARTICULIERS ORDONNÉS
IMPOSENT UN RAPPORT UNIQUE QUE L'ON PEUT DÉCOUVRIR À TRAVERS UN RÉSEAU
SEMANTIQUE

C) CERTAINS DESCRIPTEURS NE SONT PLUS AMBIGUS QUAND ON CONNAIT LEUR
CONTEXTE

CALCUL DE LA DISTANCE SEMANTIQUE ENTRE DEUX DESCRIPTEURS

**RENDRE LE SYSTEME CAPABLE DE SUGGERER ET DE CONTROLER LA
TRADUCTION**

**EXPRIMER LES LIENS SEMANTIQUES ENTRE LES DIFFERENTS DESCRIPTEURS
DU LEXIQUE**

C'EST LE ROLE DU THESAURUS

Descripteur DESC_i , Descripteur DESC_j , Relation R_k / Descripteur DESC_l

Il existe une relation R_k entre le descripteur DESC_i et le descripteur DESC_j

**le descripteur DESC_l désigne le domaine de validité en dehors duquel la relation entre les
deux descripteurs n'est plus vraie.**

LES RELATIONS

1) sorte-de

$$i \text{ sorte-de } j \iff j \text{ sorte-de }^{-1} i$$

EX: cheval **sorte-de** animal
 sysut **sorte-de** système-documentaire

2) voisin

$$i \text{ voisin } j$$

EX: bateau **voisin** navire
 voiture **voisin** automobile

3) partie-de

$$i \text{ partie-de } j \iff j \text{ partie-de }^{-1} i$$

EX: volant **partie-de** voiture
 moelle épinière **partie-de** cerveau

4) moyen

$$i \text{ moyen } j \iff j \text{ moyen }^{-1} i$$

EX: informatique **moyen** automatisation

5) associé-à

$$i \text{ associé-à } j \iff j \text{ associé-à }^{-1} i$$

EX: deuil **associé-à** noir
 médaille **associé-à** mérite

UNIVERSITE d'AIX-MARSEILLE II
 FACULTE Pierre PUGET - L.I.T.A.M.
 14, rue Puvis de Chavannes
 13001 Marseille
 Téléphone : 91 90 13 20
 Télécopie : 91 90 58 29
 Mailer EARN : KNIPPEL at FRAIX11.BITNET

VOUZZAVEDIBISAR

SEMINAIRES DE RECHERCHE 1992

GROUPE DE FORMATION DOCTORALE D'INFORMATIQUE
 FONDAMENTALE
 ET SCIENCE DE LA COMPUTATION

Salle des Séminaires Les jeudi de 14h30 à 16h

| Date | Intervenant | Thème |
|------------|---|---|
| 09/01/1992 | M.T.LASKRI Maître de Conférences Université d'Annaba | ANALYSE DE REQUETES: PASSAGE DU LANGAGE NATUREL AU LANGAGE DOCUMENTAIRE |
| 23/01/1992 | P.LIVET Professeur Université d'Aix-Marseille I | SIMULATION, MODELISATION, REPRESENTATION EN SCIENCES COGNITIVES |
| 20/02/1992 | E.BIANCO Professeur Université d'Aix-Marseille II | THEOREME DE GODEL |
| 26/03/1992 | J.B.LESSOURD Directeur de Recherches CNRS-Marseille | PROGRAMMATION LINEAIRE ET GESTION DE PORTEFEUILLES |
| 16/04/1992 | F.ADREIT Maître de Conférences Université d'Aix-Marseille III | INTERFACE HOMME-SGBD |
| 21/05/1992 | R.MARCINIAK Assistante Université d'Aix-Marseille II | GESTION DE PROJETS |
| 25/06/1992 | J.M.KNIPPEL Maître de Conférences Université d'Aix-Marseille II | THESAURUS ET RESEAUX DE PETRI |

N.B : Les textes des interventions seront disponibles au L.I.T.A.M.
 LABORATOIRE D'INFORMATIQUE THEORIQUE
 ET D'APPLICATIONS DE MARSEILLE

**Université de Provence
Atelier de Reprographie
Centre Saint Charles
3, place Victor Hugo
F - 13331 Marseille Cedex 3**