

VOLUME 27 N° 3 / MAI-JUIN 1990

**ADBS**

**DOCUMENTALISTE**

**SCIENCES DE L'INFORMATION**

ISSN 0012-4508

# VEILLE TECHNOLOGIQUE ET INFORMATION DOCUMENTAIRE

## De l'usage de la bibliométrie dans les services de documentation

Henri DOU, Parina HASSANALY, Luc QUONIAM, Albert LA TELA

*La bibliométrie, ensemble de méthodes statistiques d'évaluation de l'information scientifique et technique, entre de façon déterminante dans les systèmes d'aide à la décision tels que la veille technologique. Elle permet une gestion dynamique des informations, donnant aux services documentaires les moyens de faire face aux mutations qui ont affecté le domaine de la documentation comme la transformation qui s'ensuit des fonctions et des métiers du traitement de l'information. Devant cette situation mouvante en évolution rapide, on peut en effet s'interroger sur la place des services documentaires : comment ne pas être marginalisé dans cette nouvelle distribution des tâches, comment se former, que faire, qui satisfaire ?*

Information stratégique, surveillance sectorielle, veille technologique ou technique, information critique, système d'information, bibliométrie, scientométrie, infographie... nous nous sommes progressivement éloignés du vocabulaire classique décrivant le service de documentation. Cette évolution linguistique souligne directement la mutation qui est en train de se produire dans le domaine documentaire.

Nous allons essayer d'analyser ces changements, en redéfinissant, dans la mesure du possible, la place que

Les auteurs de cet article appartiennent au Centre de recherches rétrospectives de Marseille, Université Aix-Marseille II, 13397 Marseille Cedex 13, tél. 91.02.90.94. Henri Dou est par ailleurs président de la Société française de bibliométrie appliquée.

pourrait occuper le service documentaire dans cette nouvelle répartition des tâches. Il est bien entendu que ce

service devra de toute façon évoluer sur le plan technique pour s'insérer dans les dispositifs de demain.

## MUTATIONS ET REDÉFINITION DES FONCTIONS

### *Nouvelles technologies*

Les nouvelles technologies bouleversent rapidement les habitudes de traitement des informations. Les capacités de stockage augmentent, des facilités de plus en plus larges permettent un téléchargement ou des transferts rapides, des analyses auto-

matiques peuvent être pratiquées sur des corpus de plus en plus grands avec intégration des références internes, des stations de travail permettent la présentation de résultats sous des formes nouvelles, agréables, propices à la réflexion et à la prise de décision.

Mais il ne faut pas trouver dans ces techniques un apex qui ne consiste-

rait qu'en une manipulation étroite de ces systèmes techniques. L'équation logique de recherche n'est plus qu'un moyen, non une finalité en soi. Les nouvelles technologies ne sont pour le documentaliste moderne (on parlera de plus en plus de service d'information) que des moyens lui permettant de fournir une information à plus haute valeur ajoutée. Cela signifie en fait que le documentaliste devra dominer les systèmes techniques.

Prenons un exemple : on est passé du terminal papier au modem intégré à un micro-ordinateur. Quelle est la signification de ce changement ? Le micro-ordinateur donne au documentaliste une puissance accrue sur le plan du stockage, de la création de profils utilisables à la demande, du reformatage et de la présentation des informations. Tout en restant à un stade primaire, le MS/DOS permet, par ses tris, par la création de fichiers annexes, par ses boucles, de donner, pour tout corpus de références supérieur à une centaine d'unités, des statistiques sommaires sur les langues, les types de documents, les déposants de brevets, etc.

Ces nouvelles facilités doivent donc induire dans l'exercice même de la profession des pratiques différentes : les résultats documentaires, pour ne considérer que ceux obtenus par les recherches en ligne, doivent avoir une forme différente de ceux du passé. Même si la demande n'est pas expressément formulée par les utilisateurs, la présentation, les traitements doivent être différents.

Les nouvelles technologies sont donc un moyen de produire plus et mieux, pour formaliser l'information sous des formes plus adaptées à la prise de décision. Cela demande des compétences accrues, et souvent un surcroît de travail.

### Nouvelles méthodologies

L'accroissement de la compétitivité, la rotation rapide des technologies, les menaces permanentes que font peser des concurrents potentiels non identifiés font passer la gestion des technologies dans le domaine stratégique de l'entreprise. Il y a quelques années, le cercle étroit des dirigeants ne considérait pas comme noble la gestion de ces outils technologiques, perçus comme des moyens. De nos jours, la situation change de plus en plus rapidement, puisque la guerre économique peut être considérée comme permanente.

Cette prise en compte des faits techniques et technologiques (le technico-économique et l'économique ont déjà fait le chemin il y a quelques années), a fait apparaître la nécessité

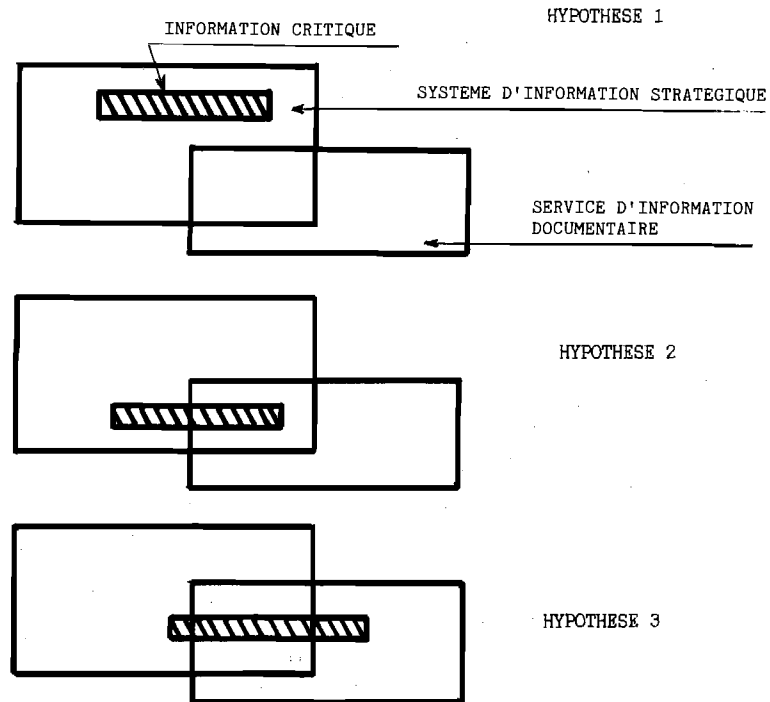


Fig. 1. - Situation du système d'information documentaire dans le système d'information stratégique de l'entreprise.

de ne plus se contenter d'informations fragmentaires, obtenues à la demande, sans concertation. Il devient nécessaire de planifier les informations à obtenir pour atteindre un but précis. Cette planification va faire apparaître des domaines nouveaux, souvent très riches, qu'il faudra prendre en compte : les informations floues comme les visites, les propos « entendus dans un colloque », les rapports de voyages, etc.

Au départ, la méthodologie d'accès à un type particulier d'information a surtout été sectorielle, ne quittant pas le service d'information documentaire qui répondait en tant que de besoin à des demandes externes : normes, recherche de brevets, etc. Mais un travail de fond, entrepris par diverses équipes de recherche et par divers praticiens, permet maintenant de dégager un axe majeur de développement : la veille technologique, qui est une surveillance sectorielle systématique, permettant par une gestion rationnelle des informations internes et externes de maîtriser les facteurs critiques de succès et de fournir des informations pour les décideurs.

La veille technologique va donc se placer directement comme un élément important de décision, au niveau du système d'information de l'entreprise. Dans ce cadre, le service d'information documentaire va être concerné par un certain nombre de tâches, passives ou actives, selon la structure de l'entreprise et sa volonté.

### Situation du service d'information documentaire

Actuellement, toutes les entreprises possèdent un système d'information stratégique. Ce dernier peut être formel ou informel, il peut concerner plus ou moins de personnes, mais rares sont les entreprises où la décision est prise sans un minimum de concertation, donc sans un minimum d'informations [1, 2, 3].

Dans ce système d'information stratégique, la veille technologique va avoir un effet structurant. En effet, elle va nécessiter la détermination de buts à atteindre, de facteurs critiques de succès à maîtriser. De ce fait, des informations critiques vont apparaître. Ce sont celles qui seront nécessaires à la prise de décision. Leur processus d'obtention, de validation, devra faire appel à des observateurs et à des experts : l'information critique a maintenant une existence réelle.

C'est par rapport à ces deux premiers systèmes cohabitants dans l'entreprise - système d'information stratégique et information critique - que devra se situer le service d'information documentaire. Il faut donc étudier comment il est actuellement perçu, et quelle est sa place. Nous avons schématisé dans la figure 1 les différentes positions respectives de ces systèmes.

On constate que le cas le plus favorable est le cas numéro 3, où un certain nombre d'informations critiques sont du ressort du service d'information

documentaire. Généralement, c'est le domaine de la propriété industrielle qui est la partie la plus concernée. Mais il y en a d'autres : les normes, la comparaison entre littérature fondamentale et brevets, l'intégration des informations floues aux informations précédentes, etc. C'est une chance, pour un service d'information documentaire, que de pouvoir se placer dans une telle configuration. Ou il saisit cette chance pour augmenter (cela nécessite souvent l'acquisition de compétences supplémentaires) la partie d'informations critiques qu'il va traiter, ou il restera passif et pourra être marginalisé.

Une chose est certaine : les informations critiques doivent être maîtrisées. Si le service d'information documentaire ne le fait pas avec suffisamment de vivacité, cette tâche lui échappera et sera effectuée ailleurs. Le service d'information documentaire doit, comme tout service, réaliser un certain nombre de tâches de fond. Mais, à côté de celles-ci, l'analyse de sa situation lui fera apparaître un certain nombre de travaux plus nobles et performants. Il lui faut concentrer son action vers ces derniers et ne pas se tromper d'objectif.

### **Les bases d'un service de veille technologique**

La veille technologique est un outil puissant qui donne aux entreprises qui la pratiquent un avantage certain en matière de prévision et de gestion de leur environnement technologique. Au départ, seule de grandes entreprises ont développé des services de ce type, souvent orientés vers la gestion des portefeuilles de brevets. L'effort en personnel nécessaire représentait un effectif de l'ordre de 5 pour 1 000... Actuellement, les entreprises d'importance moyenne veulent aussi accéder à ces modes de surveillance et de proposition. La taille ne constitue plus un handicap, car l'affinement des méthodologies, le développement d'outils automatiques, et l'utilisation optimale de l'existant sont des facteurs qui « démocratisent » la veille technologique.

La méthodologie de mise en place d'un tel service comprend les points suivants :

- Analyse de l'existant. Elle doit mettre à plat tout le système d'information de l'entreprise, au niveau formel, mais aussi au niveau informel. Si un service documentaire existe, il doit être situé dans cette démarche. En effet la veille technologique ne repose que pour partie sur les services documentaires.

- Détermination des buts de l'entreprise, des acquis à protéger à tout prix, des avancées technologiques à

maintenir, des lacunes à combler, des domaines nouveaux à approfondir. C'est de cette analyse que va découler la détermination des facteurs critiques de succès [4].

- Eclatement des facteurs critiques de succès en domaines à maîtriser sur le plan de l'information. Part de l'information ouverte, informelle, méthodologies d'acquisition et de gestion de cette information. Gestion et méthodes d'acquisition des informations extérieures floues. Cette réflexion doit tenir compte du système d'information de l'entreprise, et doit permettre de broser les grandes lignes de son évolution pour que la fonction veille technologique s'y intègre.

- Traitement des informations, mise en place de réseaux d'experts internes ou externes. La veille technologique n'est pas l'affaire de tous, mais généralement d'un certain nombre de personnes particulièrement concernées. Présentation des informations pour les cercles de décision. Il va falloir transformer l'information en renseignement !

- Evaluation des coûts. C'est à partir de celle-ci que les actions réelles seront entreprises. Cette évaluation, et le budget réel qui sera affecté à la veille, situeront le niveau des efforts possibles.

- Animation de l'activité veille technologique. Cette animation sera gérée avec soin, pour que les systèmes mis en place fonctionnent effectivement. Le service de veille technologique ne sera pas nécessairement formel, mais il existera néanmoins avec un budget.

L'examen de ces différentes modalités de mise en place met en évidence la nécessité de se reposer, pour une certaine partie des tâches de recherche, de traitement automatique et de mise en forme de l'information, sur de véritables spécialistes de l'information documentaire. Ces derniers devront bien connaître les sources et la manipulation des systèmes techniques de gestion et d'analyse des informations (en général de la micro-informatique associée à des logiciels très puissants), et si possible avoir une vision globale du système d'information de l'entreprise. Souvent, une double compétence sera nécessaire : technique, c'est-à-dire une connaissance du domaine dans lequel l'entreprise produit, et documentaire. Une triple compétence, même : la troisième, qui devient de jour en jour plus banale, concernera la maîtrise de petits systèmes informatiques, dont la puissance actuelle est bien souvent supérieure à celle des stations informatiques lourdes des années passées.

## **POUR UNE GESTION DYNAMIQUE DES INFORMATIONS : LA BIBLIOMÉTRIE**

De plus en plus, la gestion physique des informations, des livres, des rapports, ou même la réalisation répétitive de recherches en ligne ne constituent plus l'aspect principal du travail documentaire. En effet, l'analyse automatique des groupes d'informations nécessite le recours aux systèmes informatiques. Cette gestion introduit la bibliométrie, comme système d'analyse statistique dynamique permettant de positionner les sujets et les technique.

Actuellement, les théories mécaniques de la propagation des informations sont abandonnées, principalement lorsqu'il faut traiter des ensembles flous ou des informations permettant des prises de décision en milieu hostile. Ce fait rappelé, le concept de base qui sous-tend les traitements bibliométriques est le sui-

vant : soit une collection d'objets (livres, revues, publications, molécules...) contenant des informations : N<sub>1</sub>, N<sub>2</sub>, N<sub>3</sub>, ... N<sub>x</sub> ; si

$$I = \sum_{j=1}^x N_j$$

on se place dans le cas de la documentation classique, voire de l'archivage pur et simple (la lecture des informations, une par une, sans corrélation globale, entre dans ce cadre).

Si on considère qu'une information N est subdivisible en P(y) parties d'information discrètes mais significatives (fragments moléculaires, codes, thèmes de recherche, mots-clés, auteurs, institutions, pays, villes...) on peut représenter la somme globale J des informations contenues dans la collection d'objets comme :

$$I = \sum_{j=1}^x N_j = \sum_{j=1}^x \sum_{y=1}^z P_y = J \quad \left[ N_j = \sum_{y=1}^z P_y \right]$$

Poles principaux: 1 2

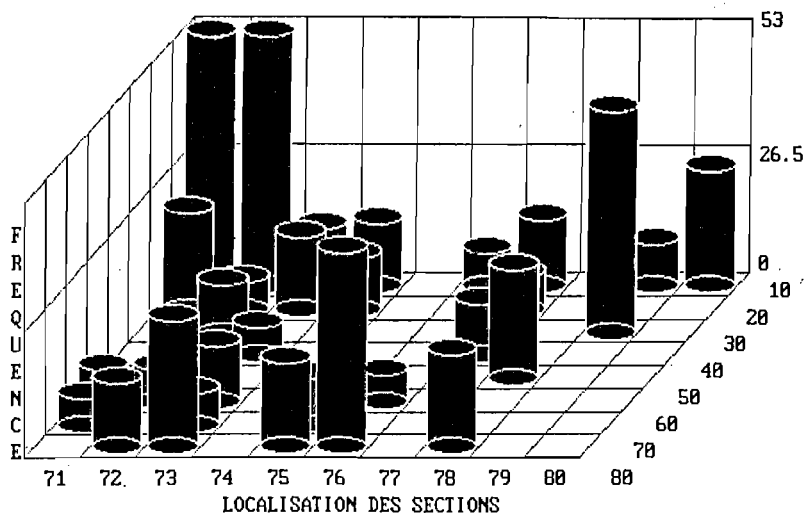
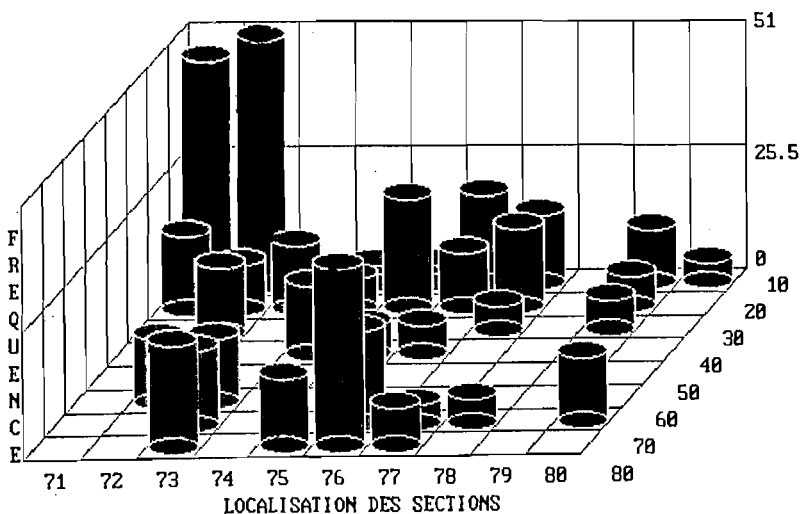
TOULOUSE 1985 - Seuil  $\geq 5$  -Pole principal 2  
2 mammalian hormonesMONTPELLIER 1985 - Seuil  $\geq 5$  -

Fig. 2. - Répartition des thèmes de travaux publiés en 1985 dans le domaine de la chimie par les universités de quatre villes françaises.

(1) Tri, représentation infographique, croisement, analyse factorielle de correspondances (AFC), analyse en composantes principales (ACP), dendrogrammes, etc.

L'AFC et l'ACP sont des traitements numériques matriciels utilisant les propriétés des espaces vectoriels euclidiens pour décrire les individus et les variables. Un dendrogramme est une représentation graphique de classification automatique. L'ensemble AFC, ACP, dendrogramme fait partie des méthodes numériques dérivées de la physique mathématique, de la théorie des communications, de la sociométrie...

(2) Codification utilisées en fonction d'un objectif précis. Exemple : classification INSPEC de la physique, classification d'appartenance politiques...

Table 1  
Section des *Chemical abstracts* et leurs codes

01	pharmacology
02	mammalian hormones
03	biochemical genetics
04	toxicology
05	agrochemical bioregulators
06	general biochemistry
74	radiation chemistry ; photochemistry ; and photographic and other reprographic processes
75	crystallography and liquid crystals
76	electric phenomena
77	magnetic phenomena
78	inorganic chemicals and reactions
79	inorganic analytical chemistry
80	organic analytical chemistry

Le concept fondamental est que, selon la combinatoire choisie (1), la somme J peut être soit identique à I, soit différente.

Ainsi, selon les taxinomies (2) mises en place (codage ou transcodage automatique) et les combinaisons choisies, on peut obtenir à partir du même corpus des informations pour l'action, appropriées aux besoins. C'est à partir de cette considération que se sont développés les systèmes d'analyses stratégiques, implémentés par des consultations d'experts, et très utilisés en veille technologique.

Les différents points examinés jusqu'ici - situation du système d'information documentaire dans une partie du système d'information de l'entreprise, mise en place d'un service de veille technologique, concepts fondamentaux des traitements modernes des informations - devraient donner au lecteur un certain nombre de pistes lui permettant de valoriser sa fonction à l'intérieur de l'entreprise. Détenteur d'un savoir, il doit éviter deux attitudes : rester passif et faire de la rétention d'information, car la connaissance n'a qu'un temps dans un domaine où tout évolue très vite.

Nous allons à présent illustrer par des exemples concrets ce que l'on peut obtenir en traitant uniquement des informations téléchargées, domaine familier au documentaliste. Ces traitements relèvent tous de la bibliométrie.

### Exemples de typologies

Cet exemple est issu du traitement automatique des données provenant des *Chemical abstracts* (connexion avec le serveur, interrogation de la banque de données, téléchargement sur ordinateur, puis traitement spécifique automatisé). La taxinomie utilisée est celle des sections des *Chemical abstracts* (80 domaines [5, 6]).

L'objectif est de montrer, par une étude des sujets des travaux publiés dans le domaine de la chimie, quelle en est la répartition par thèmes dans les principales villes françaises. La méthode est simple : on télécharge les références, on implémente la base si nécessaire, puis on traite automatiquement le champ section (DATACODE @CRRM). La figure 2 et la table 1 mettent en évidence des typologies très différentes, qui décrivent une situation globalement et sans distorsion. On peut ainsi condenser sur un même graphe une année de production, soit environ 800 à 1 000 publications.

Cette approche met en évidence des profils significativement différents. Les traitements informatiques permettent de comparer, de suivre ces productions dans le temps. L'effet des industries, des politiques... peut être mesuré en termes de production scientifique. Ceci fournit des éléments potentiellement intégrables dans un système interactif d'aide à la décision.

Dans le cas présent, les productions annuelles peuvent être intégrées à un modèle disponible sur un SGBD interne (système de gestion de base de données Texto par exemple, ou dBase III+). Ce modèle, convenablement implémenté, peut par traitement plus approfondi conduire aux réseaux de relations entre les quatre-vingts domaines, et permet de modéliser finement les interactions existantes, ou bien *a contrario* de voir pourquoi certaines n'existent pas.

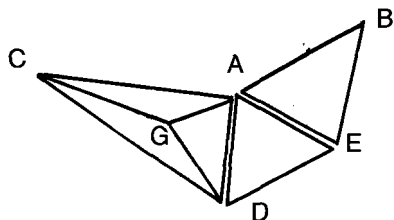
### Réseaux de relations, stratégies

Nous nous proposons ici de comparer la production scientifique des universités marseillaises dans le domaine de la physique, en utilisant la banque de données INSPEC.

On utilisera la méthode de création des réseaux à partir des paires de codes (classification INSPEC) présents dans chaque référence. Soit trois références ayant les codes suivants :

- réf. 1 : CC - A, B, E
- réf. 2 : CC - A, G, C, D
- réf. 3 : CC - D, E, A

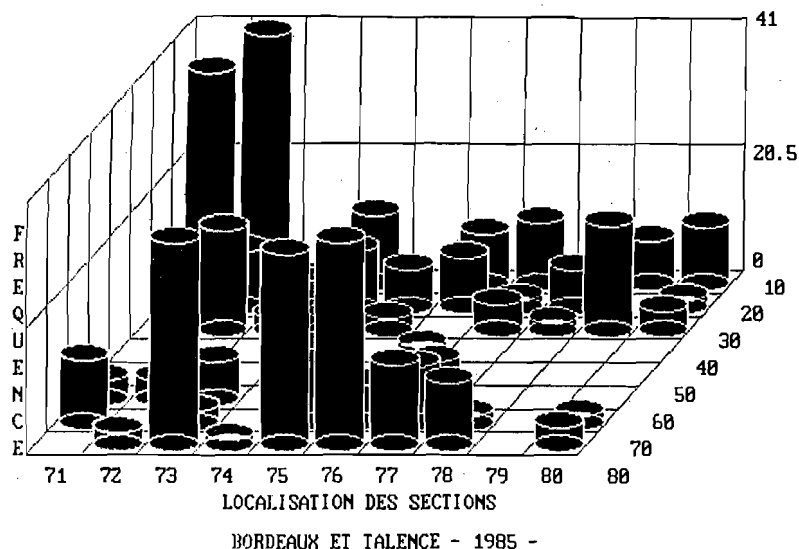
La présence simultanée dans une même référence des codes A, B, E induit des liens entre A-B, A-E, B-E. On peut ainsi établir l'ensemble des paires de codes, puis à partir de cet ensemble tracer par fréquences le réseau afférent :



La figure 3 et la table 2 mettent en évidence la différence de structure entre la production de 1984 et celle de 1988. Noter les agrégats isolés (physique théorique et astronomie), l'apparition en 1988 de travaux liés à la robotique et à l'informatique, ainsi qu'une certaine déstructuration du réseau concernant l'optique traditionnelle ou l'opto-électronique.

On remarquera que ces deux exemples - typologie et réseaux - introduisent la notion de taxinomie. Il est

Pole principal 2  
2 mammalian hormones



Pole principal 76  
76 electric phenomena

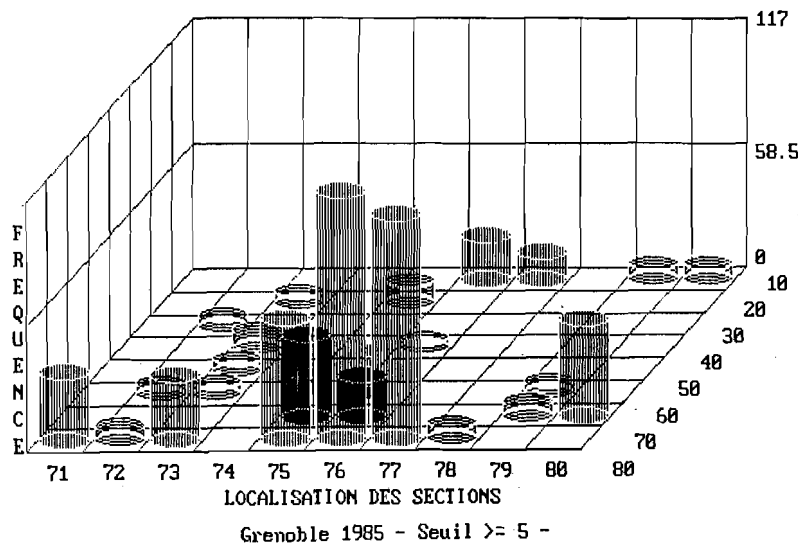


Fig. 2. - Suite.

Table 2  
Codes de la base INSPEC

001	Physics, General	A00
002	Communication, education history and phylosophy	A01
003	Mathematical methods in physics	A02
004	Classical quantum physics ; mechanics and fields	A03
005	Relativity and gravitation	A04
006	Statistical physics and thermodynamics	A05
007	Measurement science, general laboratory techn., and instrum. systems	A06
008	Specific instrumentation and techniques of general use in physics	A07
009	The Physic of Elementary particles and fields	A10
010	General theory of fields and particles	A11
011	Specific theories and interaction models ; particle systematics	A12
012	Specific reactions and phenomenology	A13
013	Properties of specific particles and resonances	A14
014	Nuclear Physics	A20
.....		
139	Software techniques and systems	C61
140	Computer Applications	C70
141	Computer applications, Administrative data processing	C71
142	Computer applications, Information science and documentation	C72
143	Computer applications, Natural sciences	C73
144	Computer applications, Engineering	C74
145	Other computer applications	C75

évident que les codes utilisés lors de l'indexation des documents ne sont pas nécessairement indiqués pour traiter les sujets liés au développement d'une entreprise. Il sera alors nécessaire de créer avec des experts internes ou externes différentes taxinomies pour l'action. Ces dernières seront utilisées pour réindexer les documents et, selon les besoins, réaliser des analyses automatiques. On peut par exemple concevoir des taxinomies politiques, ou prenant en compte des aspects techniques propres à la société, des taxinomies sur les types d'achats réalisés par les clients et sur leurs habitudes, etc. Cette utilisation de codes permet en outre une intégration facile des documents internes, des rapports, etc. Elle nécessite peu de travail de saisie. On notera au passage que l'on se situe très loin d'un thésaurus au sens documentaire du terme !

### Comparaisons entre domaines

Nous allons à présent comparer la production en chimie à Marseille à celle de l'université de Grenoble. Quels sont les thèmes majeurs communs, avec les avantages (en fréquence de publications d'une ville par rapport à l'autre), quels sont les composés chimiques utilisés uniquement à Marseille et pas à Grenoble ? Pour la période considérée (une année), le seuil de fréquence de 2 a été choisi (plus de deux publications traitant du même sujet pendant cette période).

On utilisera ici une méthode qui a été développée dans notre laboratoire. Elle utilise la puissance des traitements en ligne [7], combinée à des traitements informatiques spécifiques locaux effectués hors temps serveur (DATAGET CRRM), voir table 3.

Table 3	
<i>Méthode d'obtention et de reformatage des résultats des fichiers GET</i>	
Marseille, environ 500 références	GET IT RANK TOP 1 000
soit un GET sur les 1 000 termes différents, permet d'atteindre les fréquences 2	
Grenoble environ 1 000 références	GET IT RANK TOP 1 500
permet aussi d'atteindre les fréquences 2	
au reformatage on garde les termes ayant des fréquences entre la fréquence maximum et la fréquence 3. Soit :	
Fichier MA1 (Marseille) : 500 termes	
Fichier GRE1 (Grenoble) : 423 termes	
Noter à Marseille la dispersion plus forte pour un nombre de références plus faible.	

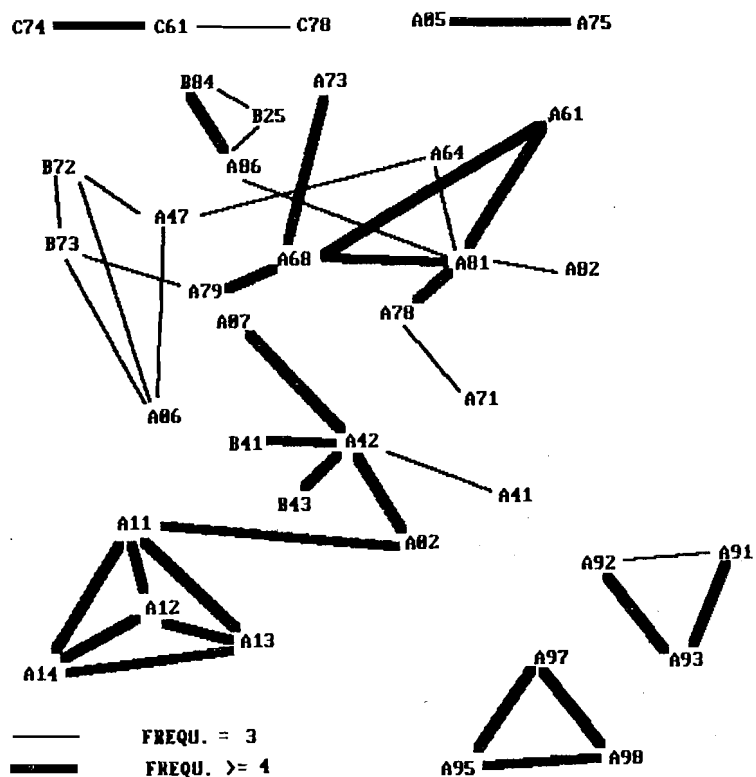
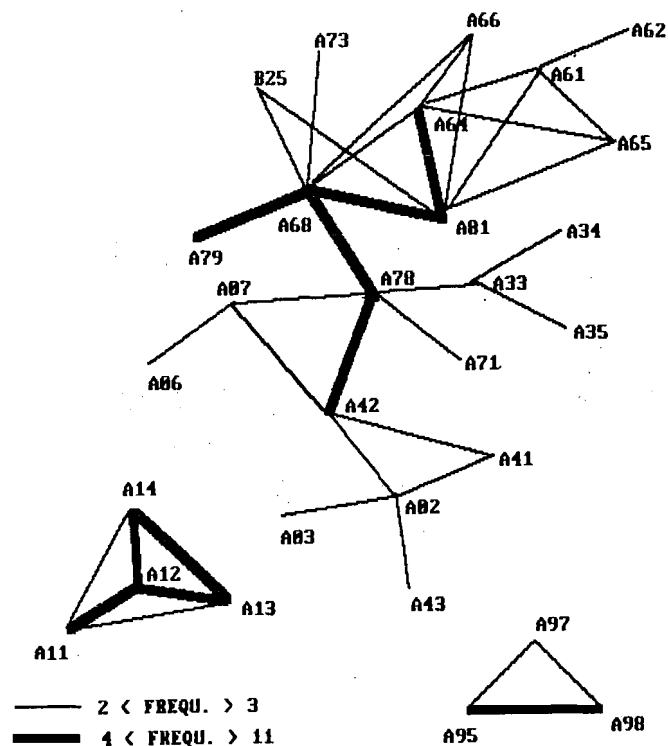


Fig. 3. - Structure par thèmes des travaux publiés en 1984 et 1988 dans le domaine de la physique par l'Université de Marseille (travaux indexés dans la banque de données INSPEC).

On sélectionne pour les deux villes (Marseille et Grenoble plus Saint-Martin-d'Hères), la production scientifique, à partir des *Chemical abstracts*. Une fois le nombre de travaux établi, on utilise en ligne la commande GET, appliquée sur le champ des *index terms* des *Chemical abstracts* (descripteurs plus numéros de registre). Cette commande permet d'éclater chaque champ des *index terms*, puis de classer l'ensemble des éléments par fréquences décroissantes. Les deux fichiers résultats (Marseille et Grenoble) sont archivés sur disque pour traitement hors temps serveur. La comparaison permet de déterminer les axes communs (on pourrait faire le contraire), ainsi que les avantages en termes de fréquences de publications. Un programme d'extraction de *registry numbers* (RN) permet de sélectionner ces derniers. Un traitement automatique des RN communs, ou différents dans les dictionnaires chimiques, permet ensuite d'accéder directement au nom chimique puis au « parent », donc aux structures de bases sur lesquelles le travail est effectué.

Une approche identique (sur la base de RN de référence, par exemple des catalyseurs) a été mise en place par nos soins dans certaines sociétés. On crée ainsi un index d'alerte permettant de signaler les variations autour des seuils critiques d'utilisation [8].

La figure 4 et les tables 4 et 5 résument partiellement les résultats obtenus.

### Les ensembles flous

L'utilisation documentaire classique des banques de données accessibles en ligne conduit à la sélection d'ensembles de références réalisés à partir de l'utilisation d'opérateurs booléens *ET, OU, SAUF* ou de leur homologues plus sophistiqués, tels les opérateurs de champs ou de proximité. Ainsi, les ensembles sélectionnés sont caractérisés par des liens directs (mots, auteurs identiques, etc.).

Dans le domaine du traitement de l'information stratégique, la sélection sur ces seuls critères est naïve. En effet, il faut, pour prendre en compte les effets de dispersion, réaliser des ensembles larges, qui seront ensuite traités informatiquement avec des outils statistiques variés. Un « bruit » contrôlé est donc nécessaire, car la question trop précise prédétermine la réponse.

Créer de nouvelles relations entre des ensembles de références est la base du système développé par E. Garfield (ISI), avec le *Science citation index*. Cette manière de travailler

GET MARSEILLE 1988  
C.A.S. CHAMP IT

GENE AND GENETIC ELEMENT	42	=====
PROTEINS	34	=====
RECEPTORS	29	=====
ANTIGENS	25	=====
AMINO ACIDS	21	=====
CONFORMATION AND CONFORME	20	=====
GENE AND GENETIC ELEMENT,	20	=====
PANCREAS	20	=====
CELL MEMBRANE	19	=====
BRAIN	18	=====
KINETICS	17	=====
PROTEIN SEQUENCES	17	=====
7440-21-3	17	=====
7439-92-1	16	=====
7440-50-8	16	=====
AMINO ACIDS, BIOL STUDY	15	=====
GEOLOGICAL SEDIMENTS	15	=====
KINETICS, ENZYMIC	15	=====
LIVER	15	=====
9001-62-1	15	=====
BIOLOGICAL TRANSPORT	14	=====
CRYSTAL STRUCTURE	14	=====
FATTY ACIDS	14	=====
GLYCOPROTEINS	14	=====
INTESTINE	14	=====
LYMPHOCYTE	14	=====
TOXINS	14	=====
ESCHERICHIA COLI	13	=====
LIPIDS	13	=====
7440-22-4	13	=====
BLOOD ANALYSIS	12	=====
FLOW	12	=====
WATERS	12	=====
12385-13-6	12	=====
56-86-0	12	=====
57-88-5	12	=====
ANTIBODIES	11	=====
CARBOHYDRATES AND SUGARS	11	=====
DIFFUSION	11	=====
EMBRYO	11	=====
GALAXIES	11	=====
OILS	11	=====
PANCREAS, COMPOSITION	11	=====
12385-13-6, PROP	11	=====
57-88-5, BIOL STUDY	11	=====
7439-89-6	11	=====
7440-44-0	11	=====
ANDROCTONUS AUSTRALIS	10	=====
CHROMATOGRAPHY	10	=====
OXIDATION	10	=====

Orbit Information Technologies, date: 04-18-1990

Fig. 4. - Représentation en histogramme d'un fichier GET (résultats de Marseille).

Table 4	
Termes présents seulement à Marseille	
25	Antigens
21	Amino acids
20	Pancreas
17	Kinetics
15	Amino acids, biol study
15	Geological sediments
15	Kinetics, enzymic
15	Liver
15	9001-62-1
14	Fatty acids
14	Intestine
14	Lymphocyte
14	Toxins
12	Blood analysis
12	Waters
12	56-86-0
12	57-88-5
11	Antibodies
11	Carbohydrates and sugars
11	Embryo
11	Galaxies
11	Oils
11	Pancreas, composition
11	57-88-5, biol study

Table 5	
Termes connus à Marseille et Grenoble	
42	Gene and genetic element
11	Gene and genetic element
34	Proteins
29	Proteins
29	Receptors
10	Receptors
20	Conformation and conformers
14	Conformation and conformers
20	Gene and genetic element, animal
04	Gene and genetic element, animal
19	Cell membrane
07	Cell membrane
18	Brain
05	Brain
17	Protein sequences
10	Protein sequences
17	7440-21-3
28	7440-21-3
16	7439-92-1
04	7439-92-1
16	7440-50-8
11	7440-50-8
14	Biological transport
06	Biological transport

permet de créer des liens à partir de la bibliographie d'une publication, permettant de dégager des ensembles d'auteurs « liés » par le même thème. Cette façon de faire, très exploitée mais pour un nombre de revues restreint, reste limitée à la littérature fondamentale.

Dans l'exemple présent, nous souhaitons déterminer les stratégies des sociétés qui dans le domaine de l'agro-alimentaire sont concernées par le PAIN; le FROMAGE, et le VIN, mais sans que ces termes soient nécessairement présents dans une même référence (ici les brevets). La figure 5 met en évidence l'impossibilité d'obtenir un tel résultat par les combinaisons booléennes, et la figure 6 indique comment la comparaison hors temps serveur des ensembles de *patent assignees* (noms des sociétés) relatifs au VIN, au PAIN et au FROMAGE (chaque mot étant pris séparément) permet de résoudre la question.

La méthodologie que nous avons développée ici est utilisée en veille technologique s'agissant des sociétés mais aussi des inventeurs (brevets), des auteurs (publications), des mots-clés, etc.

Des liens non booléens sont aussi établis lorsque la généalogie d'une recherche est effectuée par la méthode des citations, qui utilise les bibliographies des publications pour constituer des ensembles traitant des mêmes sujets sans passer par les termes d'indexation (figures 7 et 8). Cette méthode est utilisée pour déterminer des ensembles stratégiques qui seront soumis à des analyses plus poussées (par exemple l'âge des technologies).

### Gestion et connaissance des flux

Il est très important, surtout dans le domaine de l'information, de connaître les flux de publications, leur nature, leur mise en œuvre. Ceci évite, dans la politique de développement, voire d'appréciation des travaux, de commettre des erreurs dues principalement à la méconnaissance du transfert et de la diffusion des connaissances dans le domaine considéré.

Nous indiquerons ici une méthode très simple, fondée sur de simples commandes du MS/DOS : nous venons de télécharger 200 références, ce qui est courant dans certains

Fig. 5. - Recherche en ligne menée à l'aide des opérateurs booléens.

Fig. 6. - Comparaison hors temps-serveur des ensembles de noms de sociétés.

```

PROG:
ELAPSED TIME ON ORBIT: 0.02 HRS.
YOU ARE NOW CONNECTED TO THE WPI LATEST DATABASE.
COVERS BASICS 1981 - DATE: UPDATED TO
8940/UP;8940/UPEQ,8927/UPA,8912/UPB.

SS 1 /C?
USER:
~bread# and wine# and cheese#

PROG:
SS 1 PSTG (1)

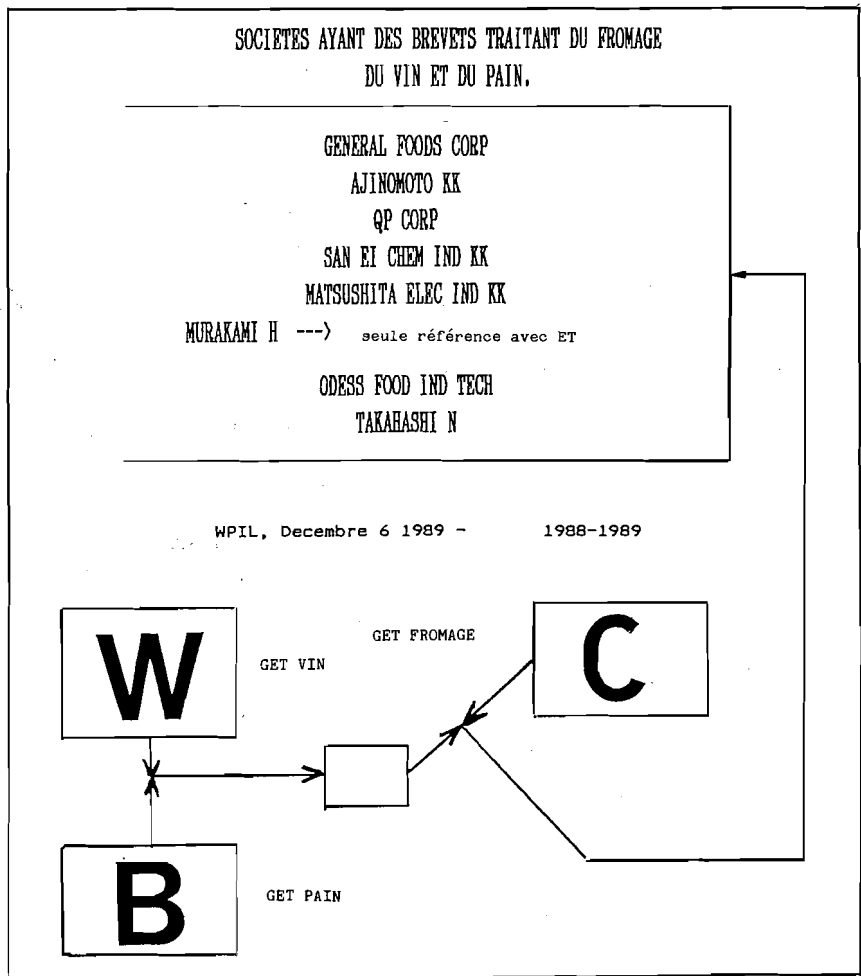
SS 2 /C?
USER:
~prt

PROG:

-1-
AN - 88-096323/14
XRAM- C88-043414
TI - Rat attractant - contains alcohol e.g. sake, wine
    etc. impregnated into food
DC - C03 P14
PA - (MURA/) MURAKAMI H
NP - 2
PN - J63048205-A 88.02.29 (8814) {JP}
    CN8704002-A 88.02.10 (8913)
PR - 86.08.19 86JP-193677 86.06.03 86JP-U84682 86.06.10
    86JP-135192 86.07.30 86JP-V17378
AP - 86.08.19 86JP-193677

SS 2 /C?
USER:
~wine# or bread# or cheese#

PROG:
SS 2 PSTG (7568)
    
```



domaines. Nous sommes donc en possession d'un fichier ASCII (nommé par exemple BIBLIO). Nous allons utiliser la commande *FIND* du MS/DOS pour avoir une vision plus nette des types de documents présents [9] :

*FIND "DT -" BIBLIO > INTER* puis *SORT <INTER> PRN* imprimera les types de document triés par ordre alphabétique indiqués dans le champ DT.

De même le champ source peut être utilisé pour traiter les principales revues, puis l'utilisation du CODE ou de l'ISSN permet d'en isoler certaines, etc. L'utilisation systématique de fichiers *batch* avec le passage de paramètres permettra, dans le cas de brevets, une observation systématique de certains *patents assignees*.

Dans cette approche, aucun logiciel particulier n'a été utilisé. Ceci prouve que même si l'on ne dispose que d'un micro-ordinateur, on peut apporter par des traitements simples des informations supplémentaires qui pourront se révéler précieuses.

### Connaissance des flux interrégionaux ou internationaux

A titre d'exemple, et toutes disciplines confondues, nous avons examiné les flux de publications de 1985 à nos jours (29/01/1990) entre Marseille, des villes françaises et certains pays étrangers. La technique de base est de rechercher les publications signées par des auteurs ayant une adresse différente, en utilisant une banque de données qui a un champ multiadresse (3) [10]. On peut aussi réaliser ce travail pour une même discipline [11].

Villes	Nb de publications communes avec Marseille
Paris	413
Montpellier	134
Toulouse	91
Strasbourg	80
Nice	49
Brest	15

Fig. 7. - Recherche effectuée par la méthode des citations. Référence issue de la banque USPA (brevets américains).

Les banques de données USPA et USPB sont accessibles via le serveur ORBIT (Orbit search service, Achilles house, Western avenue, London W3 OUA, UK). Nous remercions cet organisme pour son aide lors de la réalisation de ce travail.

(3) Ici les banques SCISEARCH (1985-1990), SCISEARCH74 (1974-1984) et SCINEW (six dernières semaines), diffusées par Orbit Search Service, Achilles House, Western Avenue, London W3 OUA, UK.

Pour les pays étrangers, le même traitement donne les résultats suivants :

Villes	Nb de publications communes avec Marseille
Etats-Unis	331
Italie	133
Espagne	67
Maroc	18
Tunisie	15
Algérie	5

Cette méthode, assez facile à mettre en œuvre, permet une meilleure connaissance des flux. Une analyse plus fine de ces chiffres met en évidence le très fort impact à Marseille du biologique et du médical, ce qui a aussi été signalé par d'autres analyses [12].

Cette utilisation met en évidence la création d'indicateurs fondés uniquement sur des fréquences, ou des rapports réalisés entre ces fréquences. Ici, on n'a même pas besoin

d'ordinateur ni de logiciel. L'accès à une base de données spécialisée est suffisant. Ceci met en relief un fait important : la réflexion est souvent plus importante que les outils.

### Techniques d'analyse, rôle des experts

Les techniques d'analyse [13, 14, 15]. Il existe sur le marché des logiciels d'analyse automatique : tris croisés, AFC, ACP... Nous ne les présenterons pas ici, car ils nécessiteraient des développements plus importants. Ils donnent en outre des informations dont l'interprétation peut rapidement devenir affaire de spécialistes. Nous conseillons donc, en un premier temps, d'utiliser des outils simples. En effet, si certains documentalistes peuvent être surpris par des traitements statistiques, que dire de l'utilisateur qui souvent n'a pas une culture documentaire très appro-

Référence issue de la Base USPA  
(brevets US)

PATENT NUMBER TITLE  INVENTOR  PATENT ASSIGNEE PUBLICATION DATE APPLICATION NUMBER PRIORITY NUMBERS/COUNTR NOTES  PATENT CLASSIFICATION  INT'L PATENT CLASS 007/00, C13K-001/00 FIELD OF SEARCH  DOCUMENT TYPE FILE SEGMENT CITATIONS(REFERENCES)	US4861381 Process for the enzymatic preparation from sucrose of a mixture of sugars having a high content of isomaltose, and products obtained Paul, Francois B., Saint-Orens-De Gammeville, FR; Monsan, Pierre F., Mondonville, FR; Remaud, Magali M. C., Ramonville, FR; Pelenc, Vincent P., Toulouse, FR Sucre Recherches et Developpements,FR 89.08.29 87.07.01 87US-068940 86.07.09 86FR-009989, FR 8609989 10 Claims, Exemplary Claim 1, 0 DRAWINGS, 0 Figures Examiner: Sneed, H. M. S. Pak, ChungK. Atty/Agent: Stevens, Davis, Miller & Mosher 127/041.000, Cross Refs: 127/046.100 X, 127/046.200 X, 435/193.000 X, 435/211.000 X, 426/658.000 X C12N-009/10, C12N-009/46, C13K- 127/046.100, 127/046.200, 127/046.300, 127/041.000, 435/193.000, 435/211.000, 426/658.000 INVENTION PATENT To Foreign Company or Corporation US2742365, 4/1956, Corman et al., 435/211. Karl L. Smiley et al., "A Simplified Method for Preparing Linear Isomalto-Oligosaccharides", Carbohydrate Research, 108 (1982) pp. 279-283. ..... The invention relates to the field of biotechnology. It relates to a process for preparing a syrup, the dry extract of which includes a high content of ..... We claim: 1. A process for preparing an aqueous syrup, the dry extract of which includes a high content of ...
--	--

LIAISON AVEC LES BREVETS

LIAISON AVEC LA LITTERATURE

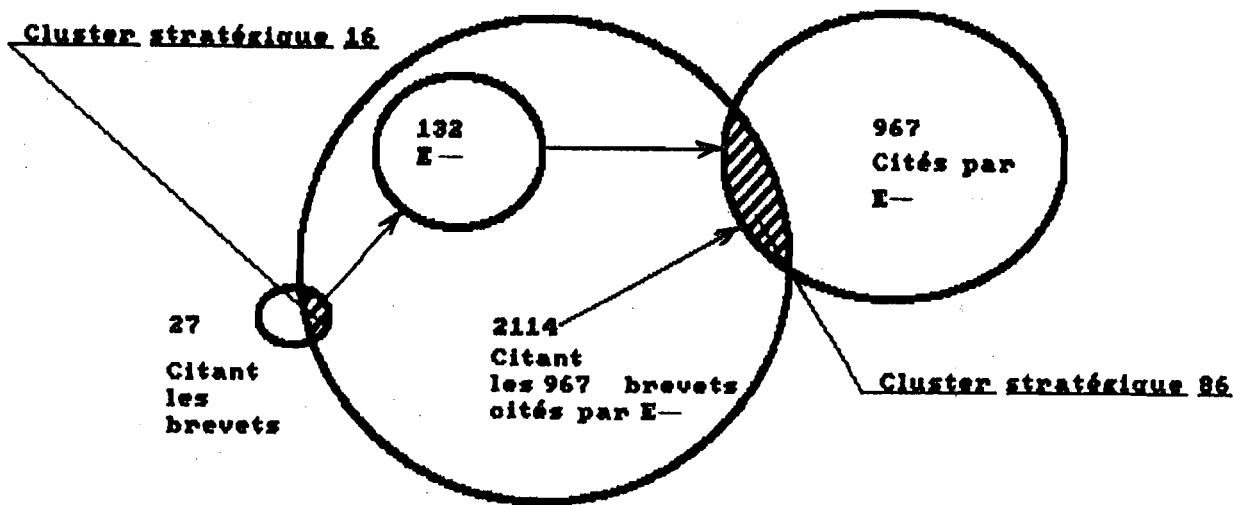


Fig. 8. - Détermination booléenne des liens stratégiques.

fondie ! Il faut essayer de donner dans ce domaine des informations fiables, ne distordant pas la vérité, et aisément compréhensibles.

Les logiciels faisant appel à des techniques élaborées de groupements de points en fonction des différents poids des informations sur les axes représentatifs sont de maniement et d'interprétation difficiles. Leur présentation nécessiterait à elle seule un article.

**Le rôle des experts.** Au-delà de l'établissement d'une simple logique d'interrogation, les experts peuvent aider les documentalistes dans leur tâche. Ils leur permettront souvent de hiérarchiser les résultats, de présenter des synthèses bibliographiques plus pertinentes. Il ne faut pas cependant penser que l'expert viendra au-devant du documentaliste, c'est ce dernier qui doit aller vers lui. Il faut donc, à une mobilité thématique, ajouter au travail documentaire une mobilité spatiale : aller vers les autres et les interroger.

\*  
\*\*

Il existe une place importante pour les services d'information documentaire dynamiques et efficaces dans le cadre des nouveaux systèmes d'information stratégique. La veille technologique, par son action structurante, et par le suivi constant de ses objectifs, va accélérer le mouvement conduisant à une nouvelle structuration des services d'information.

La cohabitation de différents spécialistes, la nécessité d'intégrer la documentation dans un ensemble plus vaste, et de ne pas attendre le questionnement systématique par autrui changera bien des habitudes. En outre, il sera nécessaire d'acquérir rapidement des compétences multiples, principalement informatiques.

L'évolution de la place du service d'information documentaire pourra se faire vers deux directions : une consultation seulement en tant que de besoin, sans participation aux prises de décisions, ou une intégration dans des systèmes dynamiques proches des décideurs, comme les systèmes

de veille technologique. Cette évolution, qui sera certainement variable selon les entreprises, sera principalement fondée sur la qualité, la compétence et la volonté des personnes concernées.

Avril 1990

## BIBLIOGRAPHIE

- [1] Maîtriser l'information critique / F. Jakobiak. - Paris : Editions d'organisation, 1988.
- [2] L'excellence technologique / J. Morin. - Paris : Publi union, Jean Picollec, 1985.
- [3] Les choix technologiques dans l'entreprise / M. Reyne. - Paris : Lavoisier, 1987.
- [4] A primer on critical success factor / C.V. Bullen, J.F. Rockart. Report CISR WP 69 SLOAN WP n° 1270-81/Center of information systems research, Sloan school of management, MIT, June 1981.
- [5] Clustering multidisciplinary chemical papers to provide new tools for research management and trends / H. Dou, P. Hassanaly, L. Quoniam, J. Kister. - *Journal of chemical information and computer sciences*, 1989, n° 29, p. 45-51.
- [6] Infographics analytical tools for decision makers. Analysis of the research production in sciences. Comparaison between Marseille and Montpellier (Chemistry) / H. Dou, P. Hassanaly, L. Quoniam. - *Scientometrics*, 1988, vol. 17, n° 1-2, p. 133-149.
- [7] GET, MAP, MEM, ZOOM et les autres / Jean-Pierre Lardy. - *Revue française de bibliométrie*, 1987, vol. 1, n° 23, p. 68-82.
- [8] Etude des Chemical abstracts : le traitement de l'IST par les indicateurs scientométriques / H. Dou, P. Hassanaly, A. La Tela, M. Millon. - *Bulletin des bibliothèques de France*, 1987, vol. 32, n° 5, p. 220-226.
- [9] Teaching bibliometric analysis and MS/DOS commands / H. Dou, L. Quoniam, P. Hassanaly. - *Education for information*, 1988, n° 6, p. 411-423.
- [10] Information systems and scientometric study in chemical oceanography / P. Hassanaly, H. Dou. - Berlin : Springer Verlag, 1986. - (Proceedings of strategies and advanced techniques for marine pollution).
- [11] European technological cooperation : the Italian participation in EUREKA / G. Napolitano. - *Science and public policy*, 1988, n° 15, p. 376-382.
- [12] Etude de la chimie à Marseille de 1981 à 1989 / H. Dou, L. Quoniam, P. Hassanaly. - *Science, technique et technologie*, 1989, n° 9, p. 34-38.
- [13] Innovation strategy of local pharmaceutical firms in Korea : a multivariate analysis / Y. Kim, L. Kim, J. Lee. - *Technology analysis & strategic management*, 1989, n° 1, p. 29-44.
- [14] A dynamic model of product and process innovation / J.M. Utterback, W.J. Abernathy. - *Omega*, 1975, n° 3, p. 639-656.
- [15] Les méthodes d'analyse des données face à l'information stratégique et l'innovation / L. Quoniam, H. Dou, C. Huot. Colloque international sur les méthodes de blocs sériation et applications, Strasbourg, 3-5 avril 1990.