

UNIVERSITE DE DROIT ET DES SCIENCES D'AIX-MARSEILLE

Faculté des Sciences et Techniques de Saint Jérôme

**Veille Technologique et Bibliométrie :
Concepts, Outils, Applications.**

THESE

Présentée et soutenue publiquement par:

ROSTAING Hervé

le 13 Janvier 1993

pour obtenir le grade de
Docteur en Sciences

Spécialité:
Sciences de l'Information et de la Communication

Président du Jury:	Rouault	(Professeur Grenoble)
Autres membres du Jury:	Carpentier	(Professeur CELSA)
	Courtial	(Professeur Nantes)
	Dou	(Professeur CRRM)
	Jakobiak	(Responsable VT à ATOCHEM)
	Quoniam	(Maître de Conférence CRRM)

Résumé en français:

Dans la compétition économique actuelle l'information est une composante essentielle de la réussite. Il est vital pour l'entreprise d'être constamment informée sur son environnement. Pour qu'une société soit compétitive elle doit faire preuve d'une forte activité innovatrice. Ce mémoire expose la réflexion menée pour concevoir et développer un logiciel d'aide à l'élaboration des stratégies des programmes en Recherche et Développement. Le premier volet traite du besoin en information scientifique et technique des entreprises, les structures et les principes à mettre en place pour une gestion efficace de cette information. Le système de Veille Technologique préconisé doit permettre l'élaboration d'indicateurs de tendances qui aideront les décideurs à diagnostiquer l'état des activités scientifiques et techniques. L'outil informatique proposé repose sur une technique de traitement automatique de l'information publiée: la bibliométrie. Après un état de l'art des méthodes bibliométriques et une analyse critique de leurs applications à la Veille Technologique, on définit un nouvel outil plus adapté à l'élaboration d'indicateurs en Veille Technologique. Ce logiciel et ses multiples traitements sont présentés ainsi que des cas pratiques d'exploitations de bases de données accessibles en ligne. La performance de cet outil repose sur ses capacités à intégrer la diversité des sources d'information, à manipuler avec souplesse les données textuelles et à livrer de multiples résultats statistiques en bibliométrie.

Résumé en anglais:

Nowadays, in the current economic competition, information is the key of success. A company needs to be continuously well informed about its environment in order to have a high innovating activity and to stay competitive. This thesis explains the thinking process which is followed in order to design and develop a software to help the elaboration of Research & Development programs. First we show industry needs in scientific and technical information. Then, we study structures and methods which must be apply to manage this information in the best possible way. Technology monitoring must allow to elaborate tendency indicators to help decision makers to diagnose the standing of scientific and technical activities. The suggested software tool is based on an automatic treatment of published information: the bibliometrics. After the review of the various bibliometrics methods and a critical analysis of their applications in strategic management of technology, we define a new method and a new tool which fits with the creation of indicators in technological management. The performance of this software results from its capacity to use various data formats, to treat these data easily and to give multiple bibliometric statistical results.

Je tiens à remercier en premier mon laboratoire d'accueil, le Centre de Recherche Rétrospective de Marseille, et plus particulièrement Monsieur le **Professeur Henri Dou** et Madame **Parina Hassanaly** qui ont su m'accorder toute leur confiance ainsi que leurs précieux enseignements.

Je tiens aussi à exprimer toute ma reconnaissance à **Luc Quoniam** sans qui mon travail ne serait ce qu'il est. J'ai trouvé en lui une compétence et un soutien de tous les jours. Je l'en remercie.

J'ai le plaisir de remercier **Albert La Tela** qui m'a apporté son expérience en informatique ainsi qu'une collaboration de tous les instants.

Je remercie **les membres du jury** et **les rapporteurs** de ma thèse de bien avoir voulu consacrer leur temps à la lecture de mon mémoire.

Je dois aussi remercier mes premières lectrices, pour leur aide méticuleuse: ma mère **Yvette** et ma compagne **Magali**. Je remercie aussi les **membres de ma famille** qui m'ont toujours accompagné durant toutes ces longues études.

Je ne pourrais terminer sans remercier **Cynthia** chez qui j'ai retrouvé la chaleur d'un foyer lorsque j'étais éloigné du mien et **William** dont la complicité dans notre travail en commun à toujours été d'une grande stimulation.

Sommaire

Sommaire

I. Introduction.....	1
II. La veille technologique et le besoin d'information élaborée en entreprise	6
A. L'incontournable besoin en information pour les entreprises.....	6
B. Le concept de veille industrielle	8
1. La notion de flux d'information	8
2. La gestion des flux d'information: la veille industrielle.....	9
C. La veille technologique: la gestion des flux d'information scientifique, technique et technologique.....	11
1. La finalité: aide à l'innovation	11
2. Les principes de fonctionnement	15
a) L'exemple du Japon.....	15
b) Méthode adaptée à la mentalité occidentale.....	16
3. La mise en application des principes	19
a) La situation courante du système d'information	19
b) Le système à préconiser	22
(1) Le renseignement systématisé.....	22
(2) L'élaboration d'indicateurs de tendance	24
(3) Le dispositif complet	28
D. La bibliométrie: technique d'élaboration d'indicateurs de tendances en veille technologique	30
1. L'aide à l'innovation.....	31
2. La source de l'information	32
3. L'aide à la prise de décision.....	33
4. La détermination des premiers facteurs critiques	35
5. Le caractère dynamique.....	37
6. La fonction des experts.....	37
7. La bibliométrie dans la veille en général.....	38
III. La source des informations exploitées en veille technologique et en bibliométrie: les bases de données.....	40
A. La distribution commerciale des bases de données	41
B. Les différents types de bases de données.....	42
1. Les bases de données scientifiques.....	45
2. Les bases de données brevets	48
3. Les bases de données de l'ISI	49
4. Les bases de données du CHI.....	52
C. Les champs indexés des bases de données.....	53

1. Les catégories d'indexations	54
2. L'enrichissement de l'indexation.....	56

IV. La bibliométrie: méthode d'évaluation des sciences et des techniques 58

A. Concept.....	58
B. Définition.....	59
C. Bref historique	61
D. La mesure de la science	64
1. Développement de la science	65
2. Le "coeur" et la "dispersion"	68
3. La modélisation des distributions bibliométriques	74
a) La loi de Bradford	75
b) La loi de Lotka	82
c) La loi de Zipf.....	85
d) Unification des lois	89
e) Mesures synthétiques des distributions	91
4. Indicateurs univariés.....	95
a) Le dénombrement des publications: indicateur de productivité	95
b) Le dénombrement des citations: Est-il un bon indicateur de qualité?	95
c) La mesure des journaux.....	97
d) La mesure des chercheurs	101
e) La mesure des laboratoires	103
f) La mesure des pays	103
g) La mesure d'un domaine.....	105
h) Classement des indicateurs bibliométriques univariés	107
5. Les cartes relationnelles	109
a) Les méthodes des co-citations	110
(1) L'association bibliographique (bibliographic coupling).....	110
(2) L'analyse de co-citation de documents.....	111
(3) L'analyse de co-citation d'auteurs	114
(4) L'analyse contextuelle des co-citations	116
(5) Critique des méthodes de co-citations	116
(6) L'analyse des citations-croisées de journaux (cross-citation).....	118
b) Les méthodes des cooccurrences de mots (co-word)	120
(1) Les réseaux "socio-techniques" (mots associés).....	120
(2) Les autres méthodes d'analyse des cooccurrences de mots	123
(3) Tableaux de contingence de mots-clés:	125
(4) Qu'apportent ces nouveaux travaux?	126
c) Les autres analyses de relations.....	129

(1) L'analyse des codes documentaires.....	130
(2) L'analyse des co-signatures.....	136
(3) L'analyse des co-opérations internationales.....	139
(4) Analyses par croisement de deux unités bibliographiques.....	141
E. La mesure des techniques et des technologies	146
1. Remise en cause des postulats bibliométriques initiaux	147
2. Indicateurs univariés.....	150
3. Les cartes relationnelles	159

V. Le logiciel bibliométrique DATAVIEW: outil d'aide à l'élaboration

d'indicateurs de tendances 164

A. Où est l'outil bibliométrique?	164
B. Caractéristiques des traitements bibliométriques.....	167
1. Diversité des sources.....	168
2. Diversité des éléments bibliométriques	169
3. Diversité des traitements bibliométriques.....	170
C. Solutions de la conception informatique adoptée	171
1. Solution à la diversité des sources.....	171
2. Solution à la diversité des éléments bibliométriques.....	177
3. Solution à la diversité des traitements ultérieurs	179
D. Description de la chaîne de traitement de DATAVIEW.....	184
1. Présentation générale des modules	184
2. Configuration de la session de travail.....	186
3. Extraction et homogénéisation des champs étudiés.....	189
4. Détermination des caractéristiques bibliométriques: le "codage"	192
5. Editions des résultats	194
a) Les statistiques des fréquences de la base	194
b) Les listes de formes.....	198
c) Les listes de paires.....	203
d) Recherche par chaînage de paires	206
e) Les tableaux	208
(1) Tableaux des fréquences de paires.....	209
(2) Tableaux binaires.....	217
(3) Tableaux d'indice d'association.....	220
E. Les traitements infographiques en sortie de DATAVIEW.....	222
1. Traitement des distributions:	224
a) Les lois bibliométriques	224
b) Listes des fréquences de formes.....	233
c) Listes des fréquences de paires.....	236

2. Traitements des tableaux:	237
a) Représentation graphique du tableau en lui-même	237
b) Les méthodes d'analyse des données	241
VI. Exemples d'études bibliométriques pouvant entrer dans un processus de Veille Technologique	246
A. Evaluation d'un secteur scientifique: étude de la production Scientifique en Chimie en France	247
B. Evaluation d'un secteur technique: Etude simultanée de trois codifications documentaires brevets	289
VII. Conclusion	315
VIII. Bibliographie	319
IX. Annexes	340
A. Annexe 1: Liste des indices d'association statistique calculés dans DATAVIEW	340
B. Annexe 2: Exemple de références bibliographiques de documents brevets.....	341
C. Annexe 3: Caractéristiques informatiques de DATAVIEW	346

Introduction

I. Introduction

Le monde de la science et de la technique se transforme rapidement et profondément, allant toujours vers plus de complexité. Les spécialités se multiplient, et les frontières qui les délimitent sont de plus en plus floues. La prise de connaissance de l'existence de ces domaines et la compréhension de leurs activités sont de plus en plus difficiles à maîtriser.

Dans ce décor changeant, **l'exploitation de méthodes, de structures et d'outils est devenue indispensable** pour mieux appréhender cette complexité. Les entreprises sont tout particulièrement concernées par ces rapides changements. Pour qu'une entreprise soit compétitive elle doit être constamment **informée des dernières découvertes, inventions ou innovations**. Elle doit pour cela s'imposer une constante observation des mutations scientifiques, techniques et technologiques.

Les entreprises, sous cet incessant besoin d'information qui les environne, mettent en place de nouvelles structures spécialisées dans la gestion de la collecte et du traitement de l'information venant de l'extérieur. Cette nouvelle activité nécessite l'application de méthodes adaptées. Des réflexions ont déjà été menées pour échafauder des méthodes de surveillance performantes et adaptées au monde industriel. Par contre, l'absence d'outils d'aide dans cette activité est un obstacle au bon fonctionnement de ces méthodes.

L'étude exposée dans ce mémoire a précisément pour objectif la conception et le développement d'un outil informatique qui répond aux exigences de cette activité de surveillance. Il est conçu pour offrir une aide considérable dans l'élaboration des stratégies des programmes de recherche et de développement.

Le premier chapitre rappelle la problématique dont souffrent les entreprises en ce qui concerne la gestion de l'information scientifique et technique. Il aborde successivement le besoin en information des entreprises, le concept de la gestion du flux d'information extérieur à l'entreprise et les systèmes de surveillance industrielle.

Puis nous déterminons l'activité de surveillance à laquelle ce mémoire se restreint: la **veille technologique**. La finalité de la veille technologique dans le processus d'innovation de l'entreprise est rappelée. On insiste tout particulièrement sur les destinataires des renseignements fournis par un service de veille technologique. Les principes de

fonctionnement et les actions à mener par ce service sont ensuite abordés. Deux principales fonctions sont préconisées dans ce mémoire:

□ **le renseignement en continu**

□ **l'élaboration d'indicateurs.**

La première fonction a déjà été étudiée par des auteurs; elle n'est que brièvement évoquée. Par contre, les modalités de réalisation de la seconde fonction sont approfondies. Une technique de traitement automatique de l'information scientifique et technique est proposée. Cette technique, basée sur le principe de la **bibliométrie**, permet de dégager des indicateurs de tendances générales concernant les activités de recherches scientifiques et techniques. La parfaite application de cette technique dans le processus de veille technologique est finalement discutée. Plusieurs raisons sont avancées pour justifier cet emploi.

Cette première partie nous permet de définir le cadre de l'étude menée lors de cette thèse ainsi que les raisons de la conception d'un outil informatique bibliométrique.

Avant d'exposer les traitements bibliométriques, un chapitre rappelle les données de base auxquelles ces traitements font appels. **Ce second chapitre fait une rapide présentation des sources d'information exploitées en bibliométrie: les bases de données accessibles par communication informatique.** La fonction de ces bases de données informatisées ainsi que les divers acteurs intervenant dans cette chaîne de diffusion d'information sont abordés. Ensuite, l'énumération des différents types de bases d'information intéressant la veille technologique est réalisée. Finalement ce chapitre se termine par une catégorie de renseignements, fournis par ces bases de données, qui est plus particulièrement utile aux traitements bibliométriques: l'ensemble des "champs indexés". **Le lecteur est mis en garde sur le fait qu'il faut une parfaite connaissance de la création de ces données lors de l'interprétation des résultats statistiques obtenus par les méthodes bibliométriques.**

Le troisième chapitre est consacré à la bibliométrie. Il présente les différentes méthodes de traitement et les grandes écoles de pensée en bibliométrie. **Ceci permet d'établir les fondements communs à toutes ces méthodes.**

En premier lieu, les postulats de départ de la bibliométrie sont énoncés. Une définition qui permet de couvrir la diversité des pratiques bibliométriques est ensuite discutée. Un bref historique est décrit, où les principaux centres français de recherche ayant une activité bibliométrique sont répertoriés.

Puis, les différentes méthodes bibliométriques ont été distribuées en deux parties distinctes selon qu'elles s'appliquent aux domaines de la science ou aux domaines de la technique et de la technologie.

La **première partie consacrée à la mesure de la science** aborde:

- les lois et les modèles mathématiques:
courbes logistiques, loi de Bradford, loi de Lotka, loi de Zipf-Mandelbrot, théorie de la communication
- les indicateurs univariés:
dénombrement des publications et des citations, indicateurs concernant les journaux, les chercheurs, les laboratoires, les pays ou un domaine
- les cartes relationnelles:
méthodes des co-citations, des citations-croisées, des cooccurrences de mots et autres méthodes d'analyse de relations

La seconde **partie consacrée à la mesure des techniques et des technologies** traite:

- tout d'abord de la remise en cause de la validité des postulats bibliométriques pour l'information technique et technologique
- des indicateurs univariés:
comptages statistiques simples et indicateurs proposés par des sociétés consultantes en stratégie technologique
- des cartes relationnelles:
rars exemples appliquant ces techniques aux informations brevets.

Ce troisième chapitre nous permet de faire l'état de l'existant en ce qui concerne les méthodes et les outils développés en bibliométrie. Il met en évidence deux points fondamentaux:

- l'inadéquation de ces méthodes aux exigences de la veille technologique**
- une absence pratiquement totale d'outils informatiques bibliométriques.**

Ces deux points sont repris dans le quatrième chapitre **pour étayer la réflexion menée lors de la conception de l'outil bibliométrique développé pendant cette thèse.** Pour que cet outil soit parfaitement adapté à l'élaboration d'indicateurs stratégiques dans le monde industriel il doit remplir trois principaux critères:

- accepter la diversité des sources d'information

- accepter la diversité des éléments bibliométriques traités
- accepter la diversité des méthodes mathématiques appliquées en bibliométrie

Le logiciel DATAVIEW, développé pendant cette thèse, a été conçu dans le respect de ces trois critères. **Ce chapitre présente les concepts bibliométriques développés pour cet outil informatique et justifie leur emploi** : notions de forme, d'occurrence de forme, de fréquence de forme, de fréquence de paire et d'indice d'association.

Ce logiciel est présenté comme une plate-forme de traitements s'insérant dans la chaîne des opérations à mener lors d'un traitement bibliométrique. Cette chaîne de traitement est décomposée pour être expliquée étape par étape. **Une partie des fonctions de DATAVIEW est explicitée** par l'intermédiaire d'exemples de traitements effectués sur un échantillon de données. Le panel des résultats bibliométriques créé par ce logiciel est présenté ainsi que des exemples de représentations infographiques. L'aspect infographique des données bibliométriques est important car il offre les avantages suivants:

- caractère synthétique des résultats
- aide à l'interprétation des résultats
- présentation adaptée à la prise de décision.

L'exposé de ce mémoire se termine par un chapitre contenant deux cas pratiques d'études bibliométriques réalisées à l'aide de l'outil bibliométrique développé dans le cadre de cette thèse.

Le premier cas met l'accent sur la nécessité de connaître parfaitement les méthodes statistiques et leur mise en application lors de traitements bibliométriques pour obtenir une interprétation exacte des résultats. D'où le besoin d'offrir des formations spécifiques au nouveaux métiers que sont ceux de la Veille Technologique et de la gestion des informations scientifiques et techniques.

La seconde étude menée en collaboration avec le CESMAP d'IBM montre l'intérêt qu'apporte l'analyse simultanée de plusieurs catégories de données bibliographiques. Une méthode d'analyse bibliométrique a été élaborée pour permettre une meilleure retranscription du contenu des innovations présentes dans les brevets. L'emploi de l'analyse relationnelle pour une étude simultanée de trois champs brevets codifiés a permis non seulement de dégager les grandes tendances pour le thème étudié, mais aussi de focaliser l'attention sur certaines innovations comportant des caractères marginaux par rapport à l'ensemble des brevets.

Pour conclure ce mémoire, nous replacerons le travail réalisé pendant cette thèse dans deux contextes. La contribution de ce travail peut s'interpréter: primo en tant que développement d'un outil d'aide dans le processus de Veille Technologique; secundo, cet outil offre aussi une plate-forme de recherche idéale pour concevoir de nouveaux traitements de données spécifiques à l'information scientifique et technique.

L'information demeure le seul bien dont la gestion est des plus problématiques car son flux est grandissant. La création d'un outil d'aide à la maîtrise de cette information doit être d'un grand soutien dans ce défi quotidien.