

UNIVERSITE DE DROIT, D'ÉCONOMIE ET DES SCIENCES
D'AIX-MARSEILLE

FACULTÉ DES SCIENCES ET TECHNIQUES DE SAINT-JÉRÔME

THESE présentée

par Charles HUOT

pour obtenir le grade de Docteur en Sciences

de l'Université de Droit, d'Économie et des Sciences d'AIX-MARSEILLE

(AIX-MARSEILLE III)

Spécialité :

Sciences de l'Information et de la Communication

ANALYSE RELATIONNELLE POUR LA VEILLE
TECHNOLOGIQUE: VERS L'ANALYSE AUTOMATIQUE DES
BASES DE DONNÉES

soutenue le 16 décembre 1992 devant la Commission d'examen:

M. Dou Henri

Professeur

M. Marcotorchino François

Professeur associé Paris VI

M. Roux Michel

Professeur Aix-Marseille II

M. Paoli Clément

Professeur associé Marne la Vallée

M. Quoniam Luc

Maître de conférence

Melle Bédécarrax Chantal

Docteur

à mes enfants Mathilde et Joseph...

Résumé de la thèse:

L'émergence de la veille technologique depuis quelques années en France, sensibilise les décideurs publics ou privés à la problématique posée par l'intelligence de leur environnement. Il y a déjà une trentaine d'années que l'on a commencé à stocker dans des bases de données des références bibliographiques relatives à des documents au sens large. Aujourd'hui ces informations correspondent à 80 % de la connaissance et elles sont accessibles à tous. Ce phénomène a donné naissance il y a une quinzaine d'années à une discipline, la bibliométrie. La lourde tâche que constitue l'analyse automatique de ces gisements de données est confiée à des outils bibliométriques fiables et performants. Les recherches présentées dans la thèse ont porté sur l'application, dans le domaine de la veille technologique, d'une méthode de classification automatique appelée l'analyse relationnelle. Cette méthode d'analyse des données opérationnelle en veille technologique, permet de pallier un certain nombre de défauts que présentaient les méthodes d'analyse employées jusqu'alors. Elle présente un caractère innovateur en rendant possible l'analyse d'informations jusqu'alors inexploitables.

MOTS CLEFS: information stratégique, analyse de données, veille technologique, analyse relationnelle, bases de données, indice de similarité, brevets, documentation automatique, analyse factorielle relationnelle, bibliométrie.

Remerciements

Avant de vous présenter ce travail, j'aimerais remercier toutes les personnes qui par leur aide ou leurs encouragements m'ont permis de réaliser cette thèse.

En tout premier lieu je tiens à remercier Chantal Bédécarrax, ingénieur de recherche au Centre Européen de Mathématiques Appliquées d'IBM. Pendant deux années Chantal a encadré cette thèse en me donnant l'occasion de collaborer à l'écriture de nombreux articles que nous avons co-signés et qui forment la majeure partie du travail présenté ici. Merci beaucoup Chantal.

Je tiens également à remercier chaleureusement François Marcotorchino directeur du Centre Scientifique d'IBM France, Henri DOU, directeur du Centre de Recherche Rétrospective de Marseille, et Parina Hassanaly responsable des enseignements du CRRM pour le sujet qu'ils ont bien voulu me confier et qui m'a passionné. pendant ces trois années.

Mes remerciements s'adressent également à Luc Quoniam, maître de conférences au CRRM de Marseille pour son soutien scientifique et moral.

Je remercie mes amis William Nivol et Hervé Rostaing, thésards en veille technologique également, avec qui j'ai eu l'occasion de collaborer à de nombreuses reprises.

Une thèse ne peut s'effectuer dans de bonnes conditions que si les équipes auxquelles on est intégré vous supportent. Ce fut le cas, et j'en remercie l'ensemble des équipes du Centre Européen de Mathématiques Appliquées d'IBM et du Centre de Recherche Rétrospective de Marseille. Merci en particulier à Pierre Michaud, Pascal Coupet, Hammou Messatfa et Albert La Teia.

Enfin, pour avoir vécu avec moi cette aventure, m'avoir soutenu de manière inconditionnelle et de plus m'avoir fait connaître les joies de la paternité pendant le temps qu'a duré cette thèse, je remercie mon épouse Anne.

Merci encore à tous.

Avant propos

Ma thèse s'est déroulée dans le cadre d'une activité industrielle au sein du Centre Européen de Mathématiques Appliquées d'IBM France. La conjoncture professionnelle dans laquelle se sont effectués mes travaux, ainsi que la mise en application immédiate lors de projets industriels concrets, m'ont poussé à avoir une réactivité forte face au processus de recherche traditionnelle, obligeant à une publication rapide des travaux réalisés.

Dans son ouvrage consacré à l'analyse de la recherche scientifique en France, [Dehe90] le professeur Paul Deheuvels de l'université Paris VI préconise de publier beaucoup et de préférence en anglais. Pour cela il s'appuie sur un slogan bien connu dans les universités américaines "*publish or perish*" (publiez ou périssez). C'est cette voie que j'ai choisie et c'est pour cette raison que ma thèse se présente sous la forme de recueil d'articles publiés dans des revues scientifiques, assorti d'une introduction générale expliquant le lien causal entre ceux-ci.

Dans un premier temps seront présentés les articles dits applicatifs dans lesquels sont reprises des études concrètes de travaux réalisés pour des clients industriels. Ces articles reproduisent à une échelle prototypale, ou sur un plan méthodologique, des réalisations qui doivent, pour des raisons de protection de l'information, demeurer confidentielles. A la fin de la thèse sont regroupés des articles plus techniques qui touchent à des développements récents de la méthode d'analyse employée. C'est au titre de l'utilisation des résultats de ces recherches dans le domaine du traitement des informations que je les ai fait figurer en annexes techniques.

Table des matières

1.0	Introduction générale	1
1.1	Introduction	2
1.2	Présentation de la thèse	3
1.3	Synthèse des articles composant la thèse	7
1.3.1	La veille technologique	7
1.3.2	Application de l'analyse relationnelle à la veille technologique: des outils d'analyse de l'information documentaire	9
1.3.3	Application d'une nouvelle méthode de classification automatique en veille technologique: l'analyse factorielle relationnelle	11
1.3.4	L'exploitation systématique des bases de données: des analyses stratégiques pour l'entreprise	13
1.3.5	A new method for analysing downloaded data for strategic decision ...	18
1.3.6	Analyse relationnelle: des outils pour la documentation automatique	21
1.3.7	Développement d'indicateurs pour l'analyse factorielle-relationnelle ...	21
1.4	Conclusion	22
2.0	La Veille Technologique	24
2.1	Introduction	25
2.2	Pourquoi la Veille Technologique?	28
2.3	La mise en place d'une structure de veille	29
2.3.1	Une Doctrine	30
2.3.2	Une Méthodologie	31
2.3.3	Une Structure	32

2.4 Les Outils de la Veille Technologique	35
2.4.1 Les Bases et les banques de données	35
2.4.2 Bibliométrie, scientométrie, infométrie	38
2.4.3 Analyse des informations	38
2.5 Conclusion	44

3.0 Application de l'A.R. à la V.T.: des outils d'analyse de l'information documentaire **45**

3.1 Introduction	47
3.2 Présentation générale des données	48
3.2.1 Extraction des références	48
3.2.2 Présentation relationnelle	49
3.3 Problèmes traités	51
3.3.1 Traitement des champs Numéros de Brevets et codes CIB	54
3.3.2 Traitement des champs Noms de Sociétés et codes CIB	57
3.3.3 Récapitulatif	60
3.4 Ouvertures	61

4.0 Application d'une nouvelle méthode de classification automatique en veille technologique: L'AFR **64**

4.1 Introduction	65
4.2 Stratégie d'interrogation	66
4.3 Analyse par les codes DERWENT	68
4.3.1 La Matrice Initiale	68
4.3.2 L'Analyse Factorielle Relationnelle	71
4.4 Annexes	81

5.0 L'exploitation systématique des bases de données: des analyses stratégiques pour l'entreprise	92
5.1 Introduction	93
5.2 Constitution du corpus de références	95
5.3 Les premiers résultats	96
5.4 Construction des tableaux à analyser	96
5.4.1 Choix des niveaux hiérarchiques	96
5.4.2 Construction des matrices de présence/absence	97
5.5 La classification des brevets	98
5.6 La classification des codes	100
5.7 Conclusion	101
<hr/>	
6.0 A new method for analysing downloaded data for strategic decision	102
6.1 Introduction	103
6.2 Basic information	104
6.3 Different ways to analyze a specific field	105
6.3.1 Frequency analysis	105
6.3.2 Pairing techniques	106
6.3.3 Matrixbuilding	107
6.4 Classical methods of data analysis	107
6.5 Relational Analysis	108
6.5.1 Data representation, basic tables	109

6.6 Relational Factorial Analysis	111
6.7 Conclusions	113
<hr/>	
7.0 Analyse Relationnelle des outils pour In documentation automatique	120
7.1 Introduction	121
7.2 Méthodologie	122
7.2.1 Classification simple	124
7.2.2 Sériation par blocs	130
7.2.3 La Quadri-Décomposition	133
7.3 Conclusion	139
<hr/>	
8.0 Développement d'indicateurs pour L'Analyse Factorielle-Relationnelle	140
8.1 Introduction	141
8.2 Développements d'indicateurs pour l'interprétation des résultats d'une Analyse Factorielle Relationnelle	141
8.2.1 Inertie totale du nuage des individus	141
8.2.2 Inertie inter-classe	142
8.2.3 Inertie between entre classes	144
8.2.4 Inertie intra classe	146
8.2.5 Développement d'indicateurs d'analyse du résultat	147
8.3 Application à la classification factorielle-relationnelle des félidés	152
8.3.1 Diagrammes factoriels et représentation des classes	154
8.3.2 Indicateurs globaux	156
8.3.3 Indicateurs par classe	157
8.3.4 Indicateurs par individu	161
8.4 Conclusion	163

8.5 Annexes **164**

9.0 Bibliographie **168**

