

UNIVERSITE DE DROIT, D'ECONOMIE, ET DES SCIENCES
D'AIX MARSEILLE
FACULTE DES SCIENCES ET TECHNIQUES DE SAINT JEROME

N° attribué par la bibliothèque

/ / / / / / / / / / / / / /

**L'HYPERTEXTE COMME MODE D'EXPLOITATION DES
RESULTATS D'OUTILS ET METHODES D'ANALYSE DE
L'INFORMATION SCIENTIFIQUE ET TECHNIQUE**

THESE

pour obtenir le grade de **Docteur en Sciences**

de l'**Université de Droit, d'Economie et des Sciences d'Aix-Marseille**

Discipline : Sciences de l'information et de la Communication

présentée et soutenue publiquement par

Luc GRIVEL

le 10 janvier 2000

JURY

M. Luc Quoniam, Professeur à l'IUT Service et Communication à St Raphael, Directeur de thèse

M. Jacky Kister, Directeur de Recherche au CNRS, Co-directeur de thèse

M. Jean-Francois Marcotorchino, Directeur du Centre Européen de Mathématiques Appliquées (CEMAP) d'IBM et Professeur associé à l'Université de Marne la Vallée

M. Thierry Lafouge, Maître de Conférence à l'Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques (ENSSIB), habilité à diriger des Recherches

M. Xavier Polanco, Responsable de l'Unité Recherche et Innovation de l'Institut de l'Information Scientifique (INIST), CNRS

Remerciements

Je tiens à remercier toutes les personnes qui, par leur aide ou leurs encouragements, m'ont permis de réaliser cette thèse :

Alain Chanudet, directeur de l'Institut d'Information Scientifique et Technique (INIST-CNRS), qui m'a permis d'effectuer cette thèse dans le contexte de l'INIST, pour son soutien,

Henri Dou, responsable du Centre de recherche rétrospective (CRRM) de Université d'Aix Marseille III, pour m'avoir accueilli dans son laboratoire,

Luc Quoniam, professeur à l'IUT Service et Communication à St Raphael, qui m'a incité à effectuer cette thèse et m'a permis de la réaliser, pour son encadrement efficace et bienveillant,

Jacky Kister, Directeur de Recherche au CNRS, qui a co-dirigé cette thèse, pour son intérêt pour mes travaux, son soutien et ses encouragements,

Jean-Francois Marcotorchino, Directeur du Centre Européen de Mathématiques Appliquées (CEMAP d'IBM) et Professeur Associé, et Thierry Lafouge, Maître de Conférence à l'Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques (ENSSIB), habilité à diriger des Recherches, pour avoir accepté la charge d'évaluer ces travaux, ainsi que pour leurs remarques,

Xavier Polanco, responsable de l'Unité Recherche et Innovation à l'INIST, avec qui je collabore depuis 10 ans, pour sa confiance, son soutien scientifique et moral,

Charles Huot, Responsable du segment 'fouille de données textuelles' à IBM, pour ses conseils et remarques,

tous les membres de l'URI, et notamment Dominique Besagni, Claire Francois, Jean Royauté, sans lesquels la majeure partie de cette recherche n'aurait pu se faire,

Jacques Ducloy, responsable de la première entité de recherche à l'INIST, qui m'a communiqué son expérience de la gestion de projets, et qui m'a fait confiance dans la conduite du projet SDOC,

tous les stagiaires pour leurs développements informatiques, et notamment trois élèves-ingénieurs qui ont participé en 1995 pendant 6 mois au projet HENOCH (Charles Broussaudier, Bruno Levy, André Kaplan), dans le cadre d'un stage de l'école supérieur en informatique et automatisme de Lorraine, (ESIAL).

Mes parents, Catherine, ...

Table des matières

Préambule	viii
Liste de mes publications par ordre chronologique	xi
Chapitre 1 De l'analyse de l'information scientifique à l'hypertexte	1
1 L'analyse de l'information scientifique et technique (IST)	2
1.1 La problématique de l'analyse de l'IST et son intérêt pour un institut tel que l'INIST	2
1.2 L'infométrie : discipline carrefour pour l'analyse de l'IST	2
1.3 Une définition opérationnelle de l'analyse de l'IST	4
2 L'hypertexte et les méthodes d'analyse de l'IST	5
2.1 Naviguer dans un océan d'information	5
2.2 La génération automatique d'hypertexte et les techniques d'analyse	5
2.3 Contexte scientifique	7
2.4 La plate-forme infométrique de l'URI	8
3 Conclusion et articulation des chapitres suivants	14
4 Bibliographie	16
Chapitre 2 Bibliométrie et cartographie de l'IST par la méthode des mots associés : démarche applicative	21
Titre original : Mapping knowledge : The Use of Coword Analysis Techniques for mapping a Sociology Data File of four Publishing Countries (FRANCE, GERMANY, UK and USA)	
Publié en 1993	
1 Introduction	22
2 Method	23
2.1 Co-words analysis.	23
2.2 SDOC programmes.	23
3 Data & Bibliometric Analysis	24
3.1 Construction of the data file	24
3.2 Application of the Bradford Law	25
4 Results and Commentary	26
4.1 Cluster analysis	27
4.2 Representing Knowledge in Scatter Diagrams	29
5 Conclusion	34
6 Epilogue	35
7 Appendix	36
8 Références	38

Chapitre 3 Apports de la linguistique informatique à l'analyse de l'IST par la méthode des mots associés **40**

Titre original : Infométrie et linguistique informatique : une approche linguistico-infométrique au service de la veille scientifique et technologique.
Publié en 1995

1	Introduction	41
2	Objectifs et hypothèse	41
3	Données, instruments et techniques	42
	3.1 Données	42
	3.2 Outil infométrique	42
	3.3 Outils linguistiques	42
4	Expérimentation	43
5	Discussion	46
	5.1 Variation et figement	46
	5.2 Indicateurs de variation et de figement	47
	5.3 Application	48
	5.4 Les clusters et les phénomènes de variation et de figement	49
	5.5 Analyse de deux thèmes représentatifs de la variation et du figement	52
6	Conclusion	56
7	Références	57

Chapitre 4 Génération automatique d'hypertextes avec cartes thématiques : avant le World Wide Web **59**

Titre original : Thematic Mapping on Bibliographic Databases by Cluster Analysis: A Description of the SDOC Environment with SOLIS
Publié en 1995

1	Introduction	60
2	Thematic Mapping	62
	2.1 Coword Analysis	62
	2.2 SDOC's clustering process	62
	2.3 The Structure of a Cluster	64
	2.4 Constructing thematic maps	65
3	Information Analysis of the SOLIS Datafile	66
	3.1 The Indexing Vocabulary	66
	3.2 Coword Clusters as Knowledge Indicators	67
	3.3 Mapping Knowledge: A Hypertext System	68
	3.4 Analysing Cluster Relationships	71
4	Conclusion	73
5	Références	74

Chapitre 5 Démarche générale d'application de méthodes d'analyse de l'IST et d'exploitation de leurs résultats

75

Titre original : Une station de travail pour classer, cartographier et analyser l'information bibliographique dans une perspective de veille scientifique et technique

Publié en 1995

1	Introduction	76
2	Choix méthodologiques et technologiques	76
	2.1 Méthodes mises en œuvre	76
	2.2 Technologie informatique	83
	2.3 La chaîne de traitement infométrique	86
3	Analyse scientométrique des résultats	88
	3.1 Exploitation des distributions bibliométriques	88
	3.2 Exploitation des résultats des méthodes d'analyse de données	88
4	Bilan et évolutions de la station de travail	101
5	Références	103

Chapitre 6 Assister l'analyse de l'IST par la génération automatique d'hypertextes dynamiques à l'ère d'internet et du World Wide Web : conception et développement d'un système d'information pour rassembler, organiser et exploiter sur INTERNET les résultats de méthodes d'analyse appliquées à des données bibliographiques

Publié en 1997

105

Titre original : A Computer System for Big Scientometrics at the Age of the World Wide Web.

1	Introduction	106
2	HENOCH system	107
	2.1 Database system	107
	2.2 Hypertexte system	107
3	HENOCH SOFTWARE CHARACTERISTICS: A GENERIC ENVIRONMENT	108
	3.1 Conversion of SGML documents into database tables	109
	3.2 A generic and extensible WWW-RDBMS gateway	110
	3.3 About HENOCH software components	111
4	AN EXAMPLE OF INFORMATION ANALYSIS ENVIRONMENT	111
	4.1 Relational modeling of informetric data	111
	4.2 Hypertext interface	112
5	Conclusion	114
6	Références	115
7	Notes	116

Chapitre 7 La conception de bases infométriques

119

Titre original : La conception de bases de données infométriques hybrides : analyse de la pratique de trois observatoires européens et proposition d'une méthode d'intégration de données hétérogènes

Publié en 1999

1	Introduction	120
2	Bases de données infométriques	121
	2.1 Présentation des organismes et de leurs objectifs	121
	2.2 Données et structure de données dans les bases infométriques	122
	2.3 Modélisation et stockage des données infométriques	
	2.4 Conclusion	128
3	Intégration de données hétérogènes	130
	3.1 Introduction	130
	3.2 Structure de données, normalisation et modèle de données : une approche intégrée pour résoudre les problèmes d'hétérogénéité des données et des formats	131
	3.3 Evaluation	132
4	Conclusion	134
5	Références	135
6	Annexes	137

Chapitre 8 Analyse de l'IST sous HENOCH : une illustration dans le domaine des plantes transgéniques

143

Titre original : HENOCH, un outil d'analyse de corpus d'information scientifique et technique

Publié en 1999

1	Présentation générale d'HENOCH	144
	1.1 A qui s'adresse HENOCH ?	144
	1.2 Qu'est ce qu'une base de données infométriques, à quoi ça sert ?	144
	1.3 Architecture informatique	145
2	Comment HENOCH permet d'explorer et d'analyser l'information scientifique et technique sans avoir à faire l'apprentissage de commandes complexes ?	146
	2.1 Comment naviguer depuis la carte thématique ?	147
	2.2 Comment analyser la carte ?	148
	2.3 Comment observer l'organisation thématique ?	148
	2.4 Comment utiliser la description bibliographique d'un document ?	153
	2.5 Comment effectuer le positionnement d'un périodique (d'un auteur, d'une affiliation, d'un mot-clé) dans les thèmes ?	154
3	Conclusion et perspectives	157
4	Références	158

Chapitre 9 Bilan critique et perspectives	159
1 Les points forts : adaptabilité et ergonomie	160
2 Les points faibles : la détection et l'analyse des évolutions thématiques dans le temps	161
3 Perspectives	162
Chapitre 10 Bibliographie générale	164
Annexe 1 : Le Programme de Recherche Infométrie (1993)	173
Annexe 2 : Une boîte à outils pour le traitement de l'information scientifique et technique (1991)	187