

UNIVERSITÉ PARIS IV - SORBONNE
U.F.R. DE LANGUE FRANÇAISE

N° attribué par la bibliothèque
99 P A 04 0027

THESE

pour l'obtention du grade de
DOCTEUR DE L'UNIVERSITÉ PARIS IV

Discipline : Linguistique

présentée et soutenue publiquement par

Mme Bénédicte BOMMIER-PINCEMIN

le 6 avril 1999

Titre :

***Diffusion ciblée automatique d'informations :
conception et mise en œuvre d'une linguistique textuelle
pour la caractérisation des destinataires et des documents***

Directeur de thèse : M. François RASTIER

JURY	Mme	Monique	SLODZIAN
	MM.	Ioannis	KANELLOS
		Xavier	LEMESLE
		Robert	MARTIN
		Jean-Marie	PIERREL
		François	RASTIER

Deuxième édition, juin 1999
avec corrections après soutenance.

Résumé

Le serveur DECID (*Diffusion Electronique Ciblée d'Informations et de Documents*), sur l'Intranet EDF, est utilisé pour trouver les personnes les plus concernées par un document ou les experts sur un sujet. La diffusion ciblée repose sur trois principes fondateurs : (i) l'automatisation et la robustesse des traitements ; (ii) le texte, aussi bien comme mode d'interrogation du système (les documents sont soumis tels quels, sans passer par des mots clés), que comme moyen de caractériser des intérêts et compétences des personnes (les profils sont calculés à partir d'un corpus de textes) ; (iii) la base formée par les profils des destinataires potentiels : toutes les activités de l'organisme sont systématiquement prises en compte.

Dans le contexte de documents écrits à dominante scientifique et technique, quatre facettes textuelles sont définies pour guider la conception des traitements : (i) la matière linguistique du texte ; (ii) son organisation interne, structurée, close et orientée ; (iii) l'intertextualité ; (iv) le rôle constitutif des lectures et la dynamique de l'interprétation. Ce modèle motive des innovations à toutes les étapes du traitement.

Un format de codage (DTD SGML) est construit, pour être à la fois robuste et général, mais aussi enregistrer des structurations significatives.

Pour la caractérisation des textes, des unités descriptives, plus contextuelles que des mots-clés, sont définies. Elles prennent en compte la détermination du local par le global et la formation des isotopies sémantiques. Ces unités descriptives sont adaptées à la recherche en texte intégral.

Une grille d'analyse est établie et appliquée à l'interprétation des formules de calcul et de mesures sur les textes.

L'ergonomie Web est adaptée à l'affichage de textes : vue simultanément globale et focalisée, points saillants, contextualisations multiples. Une représentation différentielle de la pertinence facilite le parcours des réponses du système et leur appropriation.

Abstract

Application of text linguistics to document routing

This study is part of the DECID project (*Diffusion Electronique Ciblée d'Informations et de Documents*, « targeted electronic distribution of information and documents »). The DECID team is building a document routing system for selective dissemination of information throughout the research center of the french electric power company. The three main features of targeted distribution are : (i) automatic processing ; (ii) the use of texts both for querying (instead of keywords), and for characterizing people interests and skills (profiles are based on a corpus of research program texts) ; (iii) the set of profiles represents thoroughly the research center activities.

In the scope of technical and scientific written documents, four textual dimensions are defined. They are used as guidelines in the processing development. These dimensions are : (i) the linguistical material of text ; (ii) its internal structure ; (iii) intertextuality ; (iv) the reading and hermeneutic activity as part of the text nature. Each stage of the processing is concerned by these dimensions.

A coding frame has been developed (a SGML Document Type Definition called Corpus). It enables both robust and general encoding, and also takes into account meaningful structures.

For text analysis, descriptive units are defined which are more contextual than keywords. They take into account the preeminence of global semantics in order to interpret local phenomena, and the presence of 'isotopies' (semantical recurrences). These descriptive units fit full-text search engines.

An analysis grid is established and applied to the interpretation of measurements and formula on texts.

Web interface is adapted to texts visualization. Relevance is expressed through a dynamic tree of topics. This helps the user in perceiving the results more efficiently.

Remerciements

La thèse est une recherche personnelle, mais non solitaire... et je suis heureuse de pouvoir ici exprimer ma reconnaissance envers tous ceux qui m'ont accompagnée au long de ce chemin. *Leur confiance m'a été un soutien déterminant.*

Mes remerciements vont en premier lieu à mon Directeur de thèse, François Rastier. Sa pensée est pour moi d'une richesse inépuisable et elle a profondément inspiré ma recherche. Sa disponibilité pour m'épauler a été constante. Nos rencontres de travail m'ont apporté autant par les convictions et les repères proposés, et sur lesquels s'appuyer, que par les questions ouvertes suggérées, occasions de franchir les évidences et de découvrir des réalités sous un jour nouveau.

Je suis de même profondément reconnaissante aux membres de mon jury, pour le temps et l'attention qu'ils ont consacrés à l'examen de mes travaux, alors qu'ils sont déjà abondamment sollicités par ailleurs. Leur expertise m'est précieuse. Chacun, à la croisée de sa propre recherche et de ce travail de thèse, m'a ouvert de nouvelles perspectives, qui me redonnent du recul et vont durablement nourrir ma recherche.

Monique Slodzian m'a sensibilisée à l'ampleur de la problématique du multilinguisme, qui est d'un autre ordre de grandeur que les questions d'ouverture d'un système francophone à l'anglais. La diversité des langues et l'histoire des idées dotent alors la doctrine terminologique d'une précision extraordinaire et d'un poids scientifique majeur, quelque peu estompés dans l'élan de cette thèse vers une nouvelle perspective centrée sur le texte. Je suis également reconnaissante à Monique Slodzian d'avoir été constamment attentive à me faire part de son intérêt et de sa bienveillance. Gagés sur une prise de connaissance sérieuse et attentive de mes travaux, ils sont extrêmement encourageants et stimulants pour le jeune chercheur que je suis.

Je dois à Jean-Marie Pierrel une mise au point, tout à fait fondée, sur l'apport scientifique majeur de XML. Ses observations franches et réalistes sur les plans techniques et pratiques des choix de réalisation pour l'application DECID vont être directement utiles pour la suite de mes travaux ; et je souhaite vivement poursuivre le dialogue sur ces aspects concrets et décisifs, au regard de son expérience des systèmes de Traitement Automatique des Langues, et de son bon sens efficace.

Ioannis Kanellos me donne de relire mon travail avec une profondeur nouvelle, sous l'éclairage de la question fondamentale des rapports entre la Sémantique Interprétative de François Rastier et l'informatique. Par delà un diagnostic de la valeur de mes propos sur ce point, l'enjeu véritable est d'esquisser les grandes lignes des approches possibles, en clarifiant leur manière de mettre en relation le paradigme du calcul et ses exigences (formalisation, déterminisme, robustesse), et le paradigme de l'interprétation, incluant d'autres dimensions comme l'appropriation personnelle d'un texte, l'héritage culturel d'une tradition critique.

Je suis très reconnaissante à Robert Martin pour sa rigueur et son honnêteté scientifiques et pour l'acuité de son jugement, conduisant à clarifier la confusion que pourrait entraîner ma constante référence à la sémantique dans cette thèse. La conception classique de la sémantique repose sur la reconnaissance du contenu sémantique des mots : le sens du texte s'élabore alors sur la base des apports sémantiques des mots et des effets de sens naissant de leurs interactions. Les propositions de cette thèse inversent cette vision des choses en radicalisant des orientations « rastiériennes » : pas de recours obligé à un dictionnaire, repérage d'effets globaux à l'échelle des textes et du corpus (par des calculs). Le sens est entièrement du côté de l'utilisateur et de son interprétation du fonctionnement et des propositions du système. Il faut clairement reconnaître que, par ce biais, on contourne le modèle sémantique, qui demanderait de distiller le sens au niveau du lexique, en préalable à son utilisation comme ingrédient opératoire manipulé et composé par le mécanisme.

A EDF, je voudrais d'abord remercier Xavier Lemesle, qui a suscité cette thèse et l'a accompagnée avec une conviction et une énergie jamais démenties. Sa vision stratégique et réaliste de

la diffusion ciblée a motivé bien des débats stimulants, et a donné tout son élan au projet DECID. Cette thèse lui doit beaucoup : Xavier a été, pour ma recherche, un guide sûr, par l'acuité de ses remarques et par sa clairvoyance au moment de certains choix techniques.

Cette thèse doit énormément à chacun des membres de l'équipe DECID : Pascal Obry, qui m'a partagé sa passion communicative pour Ada, et m'a fait bénéficier de son expertise en la matière ; Laurent Luciani, collaborateur informatique tout à la fois respectueux des idées qu'on lui confie et créatif, et que je tiens à remercier pour son aide quotidienne, discrète et efficace. Ma recherche est également redevable à ceux qui ont contribué à forger la diffusion ciblée à ses débuts : Marc Lamoureux, Laurent Vavasseur, Nathalie Baraban.

Je suis également reconnaissante envers Philippe Lefèvre, mon collègue de bureau, pour son enthousiasme tout à la fois chaleureux et réfléchi, et son amitié cordiale, qui m'ont accompagnée au jour le jour.

Le cadre d'EDF a été très porteur, et je remercie les responsables du Groupe TTI et du Département SID, Jean-Luc Sanson, Arlette Ferrier, François Pourbaix, Guy-Philippe Oswald, qui m'ont donné tous les moyens de mener à bien cette recherche. La constellation de compétences présentes dans cette équipe créent un climat passionnant qui a nourri ma recherche : mille et une occasions se présentent de mieux découvrir la documentation structurée et les hypertextes, les ergonomies de pointe des nouvelles technologies de l'information, les applications informatiques de la linguistique, les statistiques textuelles, les problématiques cognitives liées à la gestion et à l'usage de l'information dans l'entreprise... Tout ceci transparait au fil des pages de ce mémoire.

Le compagnonnage au sein de l'équipe Sémantique des Textes –notamment avec Denise Malrieu, Evelyne Bourion, Nathalie Deza–, et autour de cette équipe, a contribué à élargir mon horizon sur les problématiques textuelles. La qualité des échanges et des temps de réflexion, en vis à vis ou en petit groupe, les découvertes, résultats, questionnements des uns et des autres, n'ont pu que confirmer mon goût pour la Sémantique des Textes.

Je suis profondément reconnaissante à ma famille et à ma belle-famille, pour leur aide au quotidien, leur accueil toujours réconfortant, et leur compréhension patiente pour mon implication quelquefois bien prenante dans ce travail de thèse.

Mes remerciements se terminent en évoquant celui dont je parviendrais sans doute le moins à rendre compte de l'apport essentiel : mon époux François. Il m'a été un soutien extraordinaire et démultiplié : soutien moral, organisationnel, scientifique, affectueux, ménager, logistique, technique,... etc.

Sommaire

Chapitre I - Introduction	17
A. MOTIVATIONS	21
B. POINTS FONDAMENTAUX DU SUJET : UNE LECTURE DU TITRE	22
1. La diffusion ciblée	22
2. Voie explorée : la linguistique textuelle	25
3. Application concrète : évolution des versions du système DECID	27
C. CONTEXTE INITIAL	30
1. Un pôle de recherche autour des nouvelles technologies de l'information	30
2. Historique du projet	30
3. Diagnostic en 1995 (début de la thèse)	33
D. ENJEUX	38
1. Gains pour l'application DECID	38
2. Elargissement à d'autres contextes	38
3. Contribution à la Linguistique Textuelle	39
E. ORGANISATION DES CHAPITRES DE LA THÈSE	41
Chapitre II - Définir la diffusion ciblée pour l'entreprise	43
A. CONTEXTE : POURQUOI LA DIFFUSION CIBLÉE EST-ELLE UTILE, VOIRE DEVIENT INDISPENSABLE ?	49
1. De nouveaux besoins vis-à-vis de l'information	49
2. Chaque entreprise face à l'information : actions ou/et réactions	51
3. Observations sur le terrain de la DER d'EDF : difficultés pour trouver les bons interlocuteurs	53
B. PROPOSITION : COMMENT CONCEVOIR ET METTRE EN PLACE UN SERVICE DE DIFFUSION CIBLÉE ?	65
1. Propriétés attendues et caractéristiques d'un système de diffusion ciblée	65
2. Du principe à sa mise en œuvre : des questions	74
3. Éléments pour une offre de service	84
C. APPROFONDISSEMENT ET EXTENSIONS	85
1. Déclinaison d'applications utilisant des rapprochements texte - textes	85
2. Pistes à explorer	88
Chapitre III - Un panorama des applications de documentation et d'information mettant en œuvre des profils	93
A. PRÉSENTATION	99
1. Approche choisie	99
2. Bannières de ralliement	99
B. TYPOLOGIE DES USAGES DE PROFILS	101
1. Profil thématique, sélectif : recherche continue (veille) sur un sujet, au service d'une personne ou d'un ensemble d'abonnés	101
2. Profil spécifiant : différenciation des utilisateurs	107
3. Profil caractérisant et régulateur : médiation dans un espace informationnel	112
4. Profil identificateur et discriminant : repérage d'individus dans une structure	117

C.	ASPECTS CRITIQUES DE CES APPLICATIONS ET POINTS DE DISCUSSION	120
1.	Echelle	120
2.	Durée	120
3.	Degré de réalisme	120
4.	Spécificité du corpus	120
5.	L'assimilation du profil à la personne.....	121
6.	Les exigences de la mise en œuvre de l'apprentissage et le caractère adaptatif du profil.....	121
7.	Gains réels	121
D.	LA DIFFUSION CIBLÉE INNOVE : DIX POINTS POUR LA DIFFÉRENCIER DE LA DSI.....	123
1.	Délimitation de l'espace des documents	123
2.	Délimitation de l'espace des profils.....	123
3.	Personne qui a l'initiative	123
4.	Déclenchement d'une utilisation.....	124
5.	Présomption de qualité et de pertinence de l'information.....	124
6.	Degré de spécialisation de la recherche	124
7.	Dynamique.....	125
8.	Caractère (im)prévisible de l'information.....	125
9.	Equilibre bruit / silence.....	125
10.	Possibilités de reformulation	126

Chapitre IV - Eléments pour une définition de la textualité 127

A.	MULTIPLES VUES SUR LE TEXTE : L'INVENTAIRE DU COLLECTIONNEUR.....	135
1.	Avertissement	135
2.	Du côté de l'informatique et d'autres supports d'inscription et d'enregistrement	135
3.	Une linguistique qui s'aventure hors de la phrase pour aller vers le texte	137
4.	Structure et déploiement interne	139
5.	Le texte et son entour.....	141
6.	L'homme face au texte.....	145
7.	Le texte électronique : une autre textualité ?	145
B.	PROPOSITION DE SYNTHÈSE : LES QUATRE FACETTES DU TEXTE	150
1.	Le texte dans le contexte de l'application DECID : champ d'étude	150
2.	Description des quatre facettes textuelles.....	152
C.	TEXTES ET TRAITEMENTS AUTOMATIQUES : OBSERVATIONS QUANT AU STATUT DU TEXTE DANS LES PÔLES DE RECHERCHE ACTUELS	165
1.	Linguistique	165
2.	Autour de l'informatique	168
3.	Systèmes documentaires et recherche d'information : le modèle vectoriel.....	170
4.	Lexicométrie intratextuelle : l'étude des rythmes	177
D.	RECEVOIR UN TEXTE	179
1.	Compréhension.....	179
2.	Représentation	186
3.	Interprétation : huit conceptions	194
E.	LA QUESTION DE LA PERTINENCE.....	203
1.	Les expressions de la pertinence : examen des modèles rencontrés dans les applications documentaires	203
2.	Etude pour la diffusion ciblée.....	207

Chapitre V - Constitution et codage du corpus 219

A.	LES DOCUMENTS CONSIDÉRÉS	225
1.	Dissymétrie profil / requête	225
2.	Les profils.....	225
3.	Documents soumis au système pour être confrontés aux profils.....	230

B.	LES CODAGES DE TEXTES	234
1.	Des décisions significatives et déterminantes	234
2.	Cadre formel : SGML	234
3.	Quand les données sont des textes : apports de la Text Encoding Initiative	238
4.	Autres DTD standard pour les documents textuels	244
5.	La sémantique des DTD	248
C.	FORMAT DES TEXTES POUR L'APPLICATION DECID	253
1.	Conception du modèle	253
2.	Mise en œuvre du modèle : une herméneutique pour des formats électroniques ?	261
3.	Programme de lecture d'un fichier SGML : l'approche par niveaux	266

Chapitre VI - Détermination d'unités de traitement 271

A.	REPÈRES PRÉLIMINAIRES	281
1.	Rôle des unités	281
2.	Deux approches : dotation vs érosion	283
3.	Rencontre du cercle herméneutique : quelques défis pour le calcul	285
B.	EXAMEN DE TRAITEMENTS LINGUISTIQUES AUTOMATIQUES EXISTANTS	291
1.	Orientations d'utilisation des traitements courants	291
2.	Un cas concret : raisons de l'abandon momentané de l'indexation automatique	333
C.	DEUX ÉTAPES MÉDIATRICES : CONSTRUCTION, ÉLECTION	340
1.	De la nécessité de renoncer à une extraction directe des unités pour caractériser le texte	340
2.	Statut des unités élémentaires	341
3.	Statut des unités descriptives	342
4.	De l'univers descriptif au texte : l'exigence du sur mesures	345
5.	Récapitulatif : deux étapes, trois unités	346
D.	LES UNITÉS ÉLÉMENTAIRES DE DECID	348
1.	Un découpage	348
2.	Description de l'analyseur par son comportement	350
E.	LES UNITÉS DESCRIPTIVES DE DECID	355
1.	Des unités typées	355
2.	Les unités initiales	355
3.	Les unités paradigmatiques et syntagmatiques souples	359
4.	Les <i>Communautés</i>	363
F.	DISCUSSION :	
	CONFRONTATION DE LA TYPOLOGIE DES UNITÉS DESCRIPTIVES AUX APPROCHES PRÉCÉDENTES	374
1.	L'indexation : une désignation unique et des réalités très diverses	374
2.	Les systèmes de recherche sur les documents textuels :	
	les tactiques pour passer de l'expression à l'idée	377
3.	Une lecture des opérateurs documentaires (TOPIC) comme	
	explicitation de structures linguistiques et artefacts dus à la modélisation	379
4.	Définition de contextes pour le traitement automatique	396
G.	UN CHANTIER À POURSUIVRE : LA CONSTRUCTION DES COMMUNAUTÉS À PARTIR D'UN CORPUS	401
1.	Etude critique de techniques pour le groupement de mots,	
	en vue de la construction automatique de Communautés	401
2.	Choix actuel	405

Chapitre VII - Caractérisation d'un texte dans un corpus : du quantitatif vers le qualitatif 409

A.	DÉFINIR UN CORPUS	415
1.	Une question qui resurgit dans le contexte du calcul	415
2.	Le corpus : un ensemble de textes ?	416
3.	Constitution : une typologie des corpus en présence	420

B.	SÉMANTIQUE DES TEXTES, CALCULS ET MESURES	428
1.	Des chiffres et des lettres : un mélange de genres ?	428
2.	La boîte à outils conceptuelle : un modèle général pour la comparaison des formules	431
3.	La mise en formule : choix et signification - Exemples d'analyse	436
C.	PROPOSITIONS	458
1.	Moyens	458
2.	Utilisation	464

Chapitre VIII - De la pertinence au parcours interprétatif : l'interface **467**

A.	ACCÈS SÉMANTIQUE AUX BANQUES TEXTUELLES : LE TEXTE COMME POINT D'ACCÈS À D'AUTRES TEXTES	473
1.	Les modes d'interrogation des bases textuelles	473
2.	Caractéristiques de la requête textuelle - Discussion	479
3.	Le choix du texte comme base pour la caractérisation des profils dans DECID : réactions et discussion	486
B.	UNE SESSION D'INTERROGATION DE DECID : FONCTIONNALITÉS ET INTERFACE.....	490
1.	Présentation de la démarche suivie	490
2.	Accès	491
3.	Constitution de la requête	495
4.	Informations sur le traitement	501
5.	Communication des résultats	505
6.	Exploitation de la sélection.....	512
7.	Récapitulatifs des fonctionnalités	516

Chapitre IX - Conclusion **521**

A.	UNE RELECTURE DU CHEMIN PARCOURU.....	525
1.	La diffusion ciblée : réalisme et potentiel.....	525
2.	Adopter une perspective textuelle sans perdre de vue la valeur des techniques existantes.....	526
3.	Définir une sémantique textuelle pour la mise en œuvre informatique : une interprétation respectueuse et libre des théories existantes	528
B.	ET POUR POURSUIVRE : UN CHOIX DE PISTES ET DE QUESTIONS	532
1.	Dans la lignée des développements amorcés pour DECID : caractérisation de textes et calcul de liens texte - texte.....	532
2.	Aspects du traitement des textes électroniques.....	533

Note sur la numérotation des annexes

Le premier chiffre (romain) renvoie au chapitre de la thèse auquel l'annexe est plus particulièrement associée.
Le second chiffre (arabe) ordonne les différentes annexes se rapportant à un même chapitre.

Annexe II.1 - DECID en exploitation : quelques chiffres **535**

1.	En résumé	539
2.	Détail et progression chronologique	539
3.	Bilan	541

Annexe III.1 - Quelques notes à propos de TREC **543**

1.	Cadre	547
2.	Description de la tâche	547

3.	Résultats	547
4.	Analyse des conditions de l'expérience	548
5.	Variantes des tâches.....	550
6.	Eléments qui peuvent être exploités dans chacune des tâches : dissymétrie	550
7.	Données de TREC	551

**Annexe V.1 - Etude descriptive de la structure des textes d'Action (1991-1996) –
Application à l'édition électronique et au traitement automatique 553**

A.	INTRODUCTION.....	557
B.	PRÉSENTATION DU CORPUS DES TEXTES D'ACTION	558
1.	L'ordonnancement DER.....	558
2.	La forme d'une fiche d'Action, et sa version électronique	559
3.	Les parties du texte d'Action : référentiel pour cette étude	561
4.	Le corpus de travail	562
C.	POINTS DE VUE SUR LA STRUCTURATION DU TEXTE DES ACTIONS : EXISTANT.....	563
1.	Vue normative : plan-type de référence	563
2.	Vue technique : modèle de document SGML.....	563
3.	Vue algorithmique : automate de reconnaissance des titres de parties	564
4.	Vue théorique (linguistique computationnelle) : l'analyse de l'organisation générale des textes d'Action	565
D.	ELÉMENTS D'ANALYSE DE L'ORGANISATION DES TEXTES : FONCTIONS DES PARTIES.....	569
1.	Contexte.....	569
2.	But	570
3.	Etat d'avancement	571
4.	Programme (et étapes)	571
5.	Objectifs ultérieurs	573
6.	Documents de référence	574
7.	Partenariats extérieurs.....	574
8.	Autres	575
E.	VERS UNE RECONNAISSANCE AUTOMATIQUE DE L'ORGANISATION SÉMANTIQUE DES TEXTES D'ACTION	577
1.	Un point d'appui : les intitulés.....	577
2.	Les rapports entre parties.....	583
3.	Les indices linguistiques dans le contenu, au fil du texte.....	585
4.	Intégrer les indicateurs locaux et globaux.....	586
F.	PERSPECTIVES	588
1.	Utilisation de la structure des Actions	588
2.	Portée théorique et caractère généralisable de l'étude	589
G.	ANNEXES	593
1.	L'algorithme de repérage des intertitres pour le Livre Electronique	593
2.	Le vocabulaire des intitulés des parties : proposition de définition d'unités paradigmatiques	595

Annexe V.2 - La DTD Corpus 601

**Annexe V.3 - Programme d'interprétation de fichiers
et de traduction dans la DTD Corpus 605**

A.	EXEMPLE DE TRAITEMENT.....	609
1.	Comportement du programme : quelques points particuliers.....	609
2.	Une page du site du W3C	609
B.	ARCHITECTURE DU PROGRAMME ET DESCRIPTIF DES MODULES.....	615
1.	Programmes de traitement	615
2.	Les chaînes de caractères.....	615
3.	Les segments.....	616

4. Utilitaires.....	618
5. Données auxiliaires.....	619
C. CODE SOURCE.....	620
1. Programmes de traitement.....	620
2. Les chaînes de caractères.....	625
3. Les segments.....	636
4. Utilitaires.....	672
5. Données auxiliaires.....	675

Annexe VI.1 - Classification multiclasse non exhaustive –Etude générale, proposition fondée sur les Nuées Dynamiques, réalisation d’un logiciel 677

A. EXPOSÉ DE LA PROBLÉMATIQUE.....	683
1. Contexte initiateur.....	683
2. Les algorithmes existants.....	683
3. Les Nuées Dynamiques.....	684
4. Mention des méthodes alternatives envisagées les plus intéressantes.....	687
B. PRÉSENTATION DU PROGRAMME.....	690
1. Conventions.....	690
2. Entrées / sorties.....	690
3. Disponibilité.....	693
C. ALGORITHME.....	694
1. Ressources.....	694
2. Traitement.....	695
D. DISCUSSION.....	700
1. Explication des apports à l’algorithme original.....	700
2. Points d’évolution.....	703
E. APPENDICE.....	709
1. Préparation des données : choix de l’indice de dissimilarité.....	709
F. REPÈRES BIBLIOGRAPHIQUES.....	711
1. Article de référence sur la méthode.....	711
2. Présentation et commentaires concernant les Nuées Dynamiques.....	711
3. Documents complémentaires : classifications en général ; prolongements possibles.....	711

Annexe VI.2 - Schéma d’ensemble d’enchaînement des modules de traitement dans DECID 713

Annexe VII.1 - Proposition et discussion de protocoles d’évaluation pour une application de diffusion ciblée –Notes de travail 717

1. Remarques : difficultés particulières.....	721
2. Voies envisageables.....	721

Annexe VIII.1 - Synthèse des messages adressés par les utilisateurs à l’administrateur de DECID 725

A. PRÉSENTATION DU CORPUS ÉTUDIÉ.....	731
B. ANALYSE THÉMATIQUE DES RETOURS UTILISATEURS.....	732
1. Portrait des utilisateurs et des usages.....	732
2. Qualités générales reconnues à l’application DECID.....	733
3. Ecarts par rapport aux systèmes de documentation et d’information classiques - Difficultés concernant la requête.....	734

4.	Points faibles de la technique de DECID.....	735
5.	Corrections à apporter au plan de l'ergonomie de l'interface.....	737
6.	Extensions demandées.....	740
7.	Renseignements.....	741
C.	RÉPONSES AUX UTILISATEURS ET PRISE EN COMPTE DES REMARQUES.....	743
1.	Renseignement particulier.....	743
2.	Signalement d'une anomalie.....	743
3.	Cas sortant du champ de DECID.....	744
4.	Evolution à travers les versions.....	745
D.	CONCLUSIONS.....	747
1.	Les informations qui sont bien « passées ».....	747
2.	Un complément de communication sur le plan technique.....	747
3.	Les apports des nouvelles versions.....	748
4.	Objectif du déploiement : le réflexe DECID.....	748
5.	Maintien du dialogue avec les utilisateurs.....	748

**Annexe VIII.2 - Enquête auprès des utilisateurs de DECID
(septembre - novembre 1997) : Un récapitulatif des réponses 749**

A.	DÉROULEMENT DE L'ENQUÊTE.....	753
1.	Définition de la démarche.....	753
2.	Textes des questionnaires envoyés.....	754
B.	LES UTILISATIONS DE DECID.....	760
1.	Réponses sur les rubriques du questionnaire.....	760
2.	Réponses supplémentaires (rubrique « autres »).....	760
C.	AVIS SUR LES AMÉLIORATIONS POSSIBLES.....	762
1.	Réponses sur les rubriques du questionnaire.....	762
2.	Réponses supplémentaires (rubrique « autres »).....	763

Annexe VIII.3 - Planches couleur 767

Bibliographie 785