

**ORGANISATION DES
CONNAISSANCES
EN VUE DE LEUR INTÉGRATION
DANS LES SYSTÈMES
DE REPRÉSENTATION ET DE
RECHERCHE D'INFORMATION**

Textes réunis par
Jacques MANIEZ
et
Widad MUSTAFA EL HADI

**TRAVAUX
&
RECHERCHES**

51640

**Jacques MANIEZ, Widad MUSTAFA EL HADI
(Eds.)**

**Organisation des connaissances
en vue de leur intégration dans
les systèmes de représentation et
de recherche d'information**

Coordination :
Conseil Scientifique
de l'Université Charles-de-Gaulle - Lille 3

collection
UL3
travaux et
recherches

Table des matières

<i>Préface</i>	15
<i>Introduction</i>	17
<i>Vers un modèle générique des systèmes de traitement de l'information</i> Jean-Guy Meunier	21
<i>Modélisation, au moyen d'un thésaurus encyclopédique et plurilingue, des connaissances présentées au cours de l'Exposition Mondiale de l'an 2000</i> Winfried Schmitz-Esser	57
<i>Bases de connaissances terminologiques : enjeux pour la consultation documentaire</i> Nathalie Aussenac-Gilles & Anne Condamines.....	71

Modélisation des connaissances

<i>Utilisation de structures conceptuelles complexes pour représenter le « contenu sémantique » de textes narratifs</i> Gian Piero Zarri	89
<i>Extraction et modélisation des connaissances : une approche et ses technologies (EMCAT)</i> Xavier Polanco	101
<i>Intégration de connaissances pour un Système de Consultation de Documentation Technique</i> Cécile Gros & Houssem Assadi	117

Aspects linguistiques

X <i>Accès à l'information à travers les graphes de termes</i> Jean-Luc Simoni & Christian Fluhr.....	133
<i>Utilisation d'un réseau de co-occurrences lexicales pour améliorer une analyse thématique fondée sur la distribution des mots</i> Olivier Ferret, Brigitte Grau & Nicolas Masson	151
<i>Primitives sémantiques, classification des verbes et polysémie</i> Karim Chibout & Anne Vilnat	163

Aspects linguistiques : modèle et mise en œuvre

<i>Effet de l'opposition « nom plein vs anaphore » sur le traitement des textes épistémiques</i> Dominique-Guy Brassart	177
<i>Proposition de mesure de puissance sémantique</i> Pierre Maranda	191

<i>Identification de différentes structures temporelles dans des textes et leurs rôles dans le raisonnement temporel</i> Nikolay Vazov.....	199
--	-----

X <i>Apport de la statistique lexicale dans la recherche d'information</i> Patrice Bonhomme, Laurent Romary.....	213
---	-----

X <i>Impact de l'utilisation des multi termes sur la qualité des réponses d'un système de recherche d'information à indexation automatique</i> Jean-Pierre Chevallet, Marie-France Bruandet	223
--	-----

<i>Étude de l'activité des utilisateurs et repérage d'indices linguistiques pour la recherche d'information textuelle dans les documents techniques</i> Céline Paganelli	239
---	-----

Iconicité et multimédia

<i>Alex, la machine où parler, c'est montrer des représentations</i> Maryvonne Abraham	251
---	-----

<i>Nouvelle approche de conception de systèmes d'information multimédias centrée sur la distinction entre les données conceptuelles et les données sensibles</i> Isabelle Vidalenc, Laid Bouzidi	265
---	-----

<i>Modèle d'analyse pour une classification du document iconographique</i> Louis Vieira	275
--	-----

Conception de systèmes d'information

<i>Conception de systèmes d'information utilisés dans le cours d'une activité : méthodologie et outils de formalisation pour l'identification des opportunités informationnelles</i> Christian Cote	287
--	-----

X <i>L'organisation des connaissances dans les bibliothèques universitaires en France</i> Ihadjadene Majid	299
---	-----

X <i>Modélisation de l'utilisateur fondée sur ses croyances et ses buts en vue d'améliorer l'efficacité des systèmes de recherche d'information</i> Faouzi Tchenar	311
---	-----

Structuration et classification des connaissances

X <i>Vers un serveur d'indexation et de recherche sur le WEB</i> Gilbert Eymard, Jean-Marc Francony	325
--	-----

X <i>Les thesauri multilingues : interopérabilité linguistique et interopérabilité sémantique</i> Roland Dachelet	337
--	-----

Modélisation des utilisateurs. La Recherche en Science de l'information

<i>Modélisation de l'utilisateur et recherche coopérative dans les systèmes de recherche d'informations</i> Amos David	341
✕ <i>Organisation des connaissances et expertises chez les usagers</i> Elisabeth Kolmayer.....	355
✕ <i>L'organisation des connaissances en France : état des lieux</i> Yolla Polity	367

Communications affichées

<i>Une nouvelle source d'informations pour l'historien : la banque de données images de la région Nord</i> Jean-Paul Barrière, Marc de Ferrière, Marie Després-Lonnet & André Péninou	379
<i>Réorganisation de l'information scientifique en vue d'une exploitation non linéaire sous format électronique : éléments de navigation dans un document scientifique électronique</i> Mohamed Ben Romdhane & Sylvie Lainé-Cruzet	381
<i>Groupement dynamique de clients pour un système d'informations distribuées</i> Boris Charpiot, Valérie Issarny & Michel Banâtre	383
<i>Morphologie et Structuration des connaissances: pour une intégration d'outils linguistiques dans la recherche d'information en texte intégral</i> Viviane Clavier	385
<i>Quelques éléments d'algorithmique des graphes pour l'aide à la conception et à la validation de techniques d'extraction et d'organisation de connaissances langagières</i> Chantal Enguehard & Jean-Xavier Rampon	387
<i>Propositions pour une multi-indexation en philosophie</i> Benoit Hufschmitt	389
<i>Quelques mesures linguistico-mathématiques et algorithmes de filtrage pour le traitement du document écrit</i> Omar Larouk	393
<i>Discours académique disponible en Sciences Humaines : le fonctionnement discursif de l'indexation dans une base de données bibliographique automatisée</i> Clarinda Rodrigues Lucas	395
<i>Un thésaurus multilingue pour l'indexation de descriptions de marchandises : caractéristiques et liens avec la classification correspondante</i> Corentin Roulin	397

Index des sujets	401
-------------------------------	-----

Ont collaboré à ce volume

Maryvonne ABRAHAM

ENST Brest & CAMS Paris

Housseem ASSADI

LAFORIA (URA CNRS Paris 6)

Nathalie AUSSENAC-GILLES

IRIT - UMR 5505 du CNRS, Université P. Sabatier, Toulouse

Michel BANATRE

INRIA-IRISA, Rennes

Jean-Paul BARRIERE

Centre d'Histoire de la région du Nord (CHRN), Université Charles-de-Gaulle –
Lille 3, Villeneuve d'Ascq

Mohamed BEN ROMDHANE

Laboratoire RECODOC, Université Claude Bernard Lyon 1, Villeurbanne

Patrice BONHOMME

CRIN-CNRS & INRIA Lorraine, Vandœuvre-lès-Nancy

Laïd BOUZIDI

Maître de conférence I.A.E., Université Jean Moulin Lyon 3

Dominique Guy BRASSART

Théodile-Crel, Université Charles-de-Gaulle – Lille 3, Villeneuve d'Ascq

Marie-France BRUANDET

CLIPS IMAG, Grenoble

Boris CHARPIOT

INRIA-IRISA, Rennes

Jean-Pierre CHEVALLET

CLIPS IMAG, Grenoble

Karim CHIBOUT

Groupe langage et cognition, LIMSI-CNRS, Orsay

Viviane CLAVIER

CRISTAL-GRESEC, Université STENDHAL, Grenoble

Anne CONDAMINES

Equipe de recherche en syntaxe et sémantique (ERSS), UMR 5610 CNRS,
Toulouse

Christian COTE

ERSICO-Université Jean Moulin (Lyon) 3

Roland DACHELET

Equipe ATOLL-INRIA, Rocquencourt; Université Robert Schuman, Strasbourg 3

Amos DAVID

Centre de recherche en informatique de Nancy (CRIN), Vandœuvre-lès-Nancy

Marie DESPRES-LONNET

CREDO ER CNRS 128, Université Charles-de-Gaulle – Lille 3, Villeneuve d'Ascq

Chantal ENGUEHARD

Faculté des sciences et techniques de l'Université de Nantes

Gilbert EYMARD

Information-Communication, Equipe « Recherche Intelligente, Interactive d'Information » (RI3), IUT2 Université Pierre Mendès France, Grenoble

Olivier FERRET

LIMSI-CNRS, Orsay

Marc de FERRIERE

Centre d'Histoire de la région du Nord (CHRN), Université Charles-de -Gaulle – Lille 3, Villeneuve d'Ascq

Christian FLUHR

DIST/DIR, CEA/Saclay, Gif/Yvette

Jean-Marc FRANCONY

Equipe R13, Département Information Communication IUT 2, Université Pierre Mendes France, Grenoble 2

Brigitte GRAU

LIMSI-CNRS, Orsay

Cécile GROS

EDF-DER IMA/TIEM/SOAD, Clamart

Benoit HUFSCMITT

Centre de Documentation de Bibliographie Philosophique (CDBP), Section de Philosophie, Université de Franche Comté, Besançon

Madjid IHADJADENE

ENSSIB, Villeurbanne

Valérie ISSARNY

INRIA - IRISA, Rennes

Elisabeth KOLMAYER

ENSSIB-CERSI, Villeurbanne

Sylvie LAINE-CRUZEL

Laboratoire RECODOC, Villeurbanne

Omar LAROUK

CERSI-ENSSIB, Villeurbanne ; IUT-Dijon Dept. Info-Com

Clarinda Rodriguez LUCAS

Universidade estadual de campinas, Brésil

Pierre MARANDA

Université Laval, Québec

Nicolas MASSON

LIMSI-CNRS, Orsay

Jean Guy MEUNIER

Laboratoire d'Analyse Cognitive de l'Information (LANCI), Université du Québec à Montréal

Céline PAGANELLI

CRISTAL-GRESEC, Université STENDHAL, Grenoble

André PENINOÙ

Centre de Ressources Informatiques, Université Charles-de-Gaulle – Lille 3, Villeneuve d'Ascq

Xavier POLANCO

Programme de Recherche en Infométrie (PRI), Institut de l'Information Scientifique et Technique (INIST), Centre National de la Recherche Scientifique (CNRS), Vandœuvre-lès-Nancy

Yolla POLITY

RI3 (Recherche interactive et intelligente de l'information), UT2 de Grenoble, Université Pierre Mendès-France (Grenoble 2)

Jean-Xavier RAMPON

Faculté des sciences et techniques de l'Université de Nantes

Laurent ROMARY

CRIN-CNRS & INRIA Lorraine, Vandœuvre-lès-Nancy

Corentin ROULIN

BJL, Bruxelles

Winfried SCHMITZ-ESSER

Diplom-Volkswirt, Information Systems Consultancy, Hamburg

Jean-Luc SIMONI

CEA/Saclay, Gif/Yvette

Faouzi TCHENAR

CERSI DOC Lyon 1, ENSSIB, Villeurbanne

Nikolai VAZOV

Department of Russian Studies, Sofia University

Isabelle VIDALENC

ERSICO, Université Jean Moulin Lyon 3

Louis VIEIRA

Université Nancy II, IUT A – Département Information et Communication

Anne VILNAT

Groupe langage et cognition, LIMSI-CNRS, Orsay

Gian Piero ZARRI

CNRS EHESS-CAMS, Paris

PRÉFACE

Pour quelqu'un qui n'a parcouru, en curieux, que quelques-uns des sentiers déjà frayés du domaine des sciences de l'information et de la documentation, le bilan de l'exploration collective ici présenté est impressionnant et passionnant.

Impressionnant, certes, par la richesse et la variété de la collecte ainsi rassemblée, par le dynamisme et la hardiesse de ses auteurs, mais plus encore par les promesses qu'il recèle et les espérances qu'il entretient. Passionnant aussi par les nouveautés qu'il nous découvre, les curiosités qu'il nous réserve et les avancées qu'il nous ménage, nous invitant ainsi à penser qu'en ce domaine, hier est déjà entré dans la préhistoire et aujourd'hui dans l'histoire.

Mais ce qui mérite, selon nous, une attention particulière, c'est que ces pages illustrent largement et de manière éclatante la fécondité d'un paradigme épistémologique naissant, que l'on rattache aux sciences de l'artificiel et de l'ingénierie. Ce paradigme, où les sciences « molles » se confrontent aux « dures », où se rencontrent tant de disciplines relevant des unes et des autres, qu'il s'agisse des sciences cognitives, de l'intelligence artificielle, de l'informatique, des sciences de l'information et de la communication, de la linguistique, de la sémiologie et de la pragmatique, de la systémique..., se caractérise, semble-t-il, par une visée propre qui est moins de découvrir par suite d'analyses multiples les lois causales permettant d'expliquer une réalité (la forme, le sens...) que de modéliser l'ensemble des processus permettant d'évaluer un usage et d'optimiser une pratique.

C'est cette méthodologie du constructible ou du faisable qui, par la visée téléologique qui la fonde et l'anime, fait en sorte que la pluridisciplinarité qui se déploie vertigineusement dans ces pages, loin de conduire à la dissémination ou à l'éclatement de son objet, se métamorphose heureusement en transdisciplinarité féconde.

L'université de Lille III, qui s'est honorée d'accueillir en octobre dernier le chapitre français de l'ISKO, est heureuse aujourd'hui d'en publier les actes, comme elle se réjouit de retrouver en août prochain pour son colloque international les membres de cette active association.

Gérard Losfeld, Président de l'Université de Lille III.

INTRODUCTION

Les premières journées du Chapitre français de l'ISKO (International Association for Knowledge Organization) se sont déroulées dans le cadre sympathique de la maison de la recherche de l'Université de Lille 3, en présence d'une centaine de participants.

Fondée en 1989, l'ISKO a pour ambition de favoriser la recherche et les échanges entre les spécialistes de la documentation et ceux des disciplines adjacentes ou sous-jacentes telles que les sciences du langage, les sciences cognitives, l'informatique et notamment l'intelligence artificielle.

Nous pensons que cette première manifestation a été fidèle à cet esprit d'ouverture si l'on en juge par la diversité des intervenants et des interventions :

- Diversité dans l'origine géographique des intervenants : France, Canada, Allemagne, Belgique, Bulgarie, Brésil.

- Répartition assez équilibrée entre recherche théorique et recherche appliquée.

- Diversité dans les disciplines d'origine : science de l'information, sciences cognitives, informatique, linguistique, philosophie, mathématiques, intelligence artificielle... Inventaire par nature difficile, la plupart des équipes étant pluridisciplinaires. Toutefois on est impressionné par la quasi-omniprésence de l'informatique, qu'elle apparaisse sur le devant de la scène ou à l'arrière-plan.

- Diversité enfin des thèmes abordés.

Conformément à la tradition, les textes des interventions sont regroupés et classés selon l'ordre de succession des séances. Nous y avons ajouté une présentation succincte des neuf communications affichées (posters) qui n'ont pas pu être intégrées au programme chargé des exposés des deux journées.

Outre la table des matières, l'index fournit un accès circonstancié au contenu des interventions. Enfin, à l'intention des lecteurs pressés, nous

proposons ci-après un récapitulatif des thèmes abordés, en les regroupant sous trois rubriques : études de portée générale, développement des méthodes de recherche des informations, problèmes de l'image.

Études de portée générale

Épistémologie

- Modèle générique sur des systèmes de traitement de l'information (p. 21).
- Le point sur la recherche francophone des 50 dernières années en science de l'information (p. 367).

Méthodes de traitement du langage applicables à l'analyse cognitive ou à la représentation abstraite des concepts :

- constitution d'un réseau hiérarchique de primitives sémantiques des verbes français en vue d'un système expert susceptible de résoudre en contexte les ambiguïtés dues à un emploi métaphorique (p. 163)
- recherche de « lois de marquage » du discours scientifique, (signaux apparents dans la structure du paragraphe) et étude du marquage affiché par l'opposition 'mot plein/anaphore' (p. 177)
- étude de la valeur significative des suffixes lexicaux (dans les textes et dans les titres) comme marque distinctive de l'opposition OBJET/ACTION (p. 239)
- mesure statistique de la puissance sémantique des termes fortement polysémiques (p. 191)
- identification des différentes structures temporelles des textes par détection des marqueurs de temps à l'intérieur de chaque paragraphe en vue d'une typologie des raisonnements temporels (p. 199)
- outils de formalisation des processus d'une activité professionnelle en vue de leur utilisation dans un système expert (p. 287)
- création de structures conceptuelles complexes destinées à une représentation normalisée du contenu de textes narratifs (p. 89).

Analyses critiques des pratiques de l'indexation

- propositions pour une nouvelle conception du thésaurus appliquée au domaine philosophique (p. 389)
- critique de l'indexation à base de thésaurus en tant que passage obligé et déformant par une structure conceptuelle prédéfinie, avec l'exemple d'un corpus de sociologie (p. 395).

Développements et améliorations de la recherche d'informations

1. Par raffinement et extension de la structure des thésaurus

- problèmes des thésaurus plurilingues et solutions diverses permettant d'assurer la compatibilité sémantique complète entre les langues (p. 337)
- exemple d'un thésaurus encyclopédique plurilingue constituant la pièce maîtresse d'un système de gestion de connaissances en vue d'une exposition mondiale (p. 57)
- exemple d'un thésaurus plurilingue destiné à l'indexation de descriptions de marchandises, et doté de ce fait de spécificité : descripteurs concrets vs abstraits, relation qualifiant/qualifié (p. 397).

2. Par le développement de liens hypertextuels

- résultats d'une enquête sur l'efficacité des listes d'autorité (notamment Rameau) dans les catalogues informatisés des bibliothèques (OPACS) et propositions visant à améliorer la visualisation des index et le butinage (p. 299).

3. Par la prise en compte de facteurs extérieurs au contenu documentaire, notamment l'utilisateur

- modélisation de la structure physique du document : schéma, tableaux... (p. 381)
- modélisation des usagers, développée dans quatre exposés : modèle du profil (p. 383), par typologie des expertises des usagers (p. 355), par recherche coopérative (p. 341), par un système de thésaurus personnalisé (p. 311).

4. Par le raffinement des techniques d'indexation automatique

B4a - Procédés logico-mathématiques

- perfectionnement de l'analyse thématique fondée sur la distribution des mots par un réseau de cooccurrences lexicales (p. 151)
- utilisation de la technique des graphes (p. 133, p. 387)
- utilisation de l'écart réduit en statistique lexicale (p. 223)

B4b - Procédés d'analyse linguistique

- extraction de syntagmes significatifs (p. 223)
- exemple d'une base de connaissances terminologiques construite à partir d'un corpus de textes et contrôlée par des spécialistes (p. 71)
- repérage des suffixes lexicaux (p. 385)
- repérage des signes de ponctuation (p. 393)

5. Par des processus interactifs complexes

- approche conjuguant les techniques de l'infométrie, de la linguistique informatique et de l'intelligence artificielle pour une analyse complète de l'information (p. 101)
- approche spécifique des bases multimédias avec un traitement différencié des données sensibles et des données conceptuelles (p. 265)
- proposition d'une méthode d'indexation fondée sur la distinction fondamentale entre Objets et Actions (p. 239)
- exemple d'un système de recherche d'une banque documentaire technique conjuguant la complémentarité de divers instruments : table des matières, texte intégral, langage documentaire, index, modèle d'activité (p. 117).

6. La recherche sur l'Internet

- Exemple d'une banque de données sur le Web regroupant 70 sites documentaires et utilisant la classification Dewey comme langage fédérateur (p. 325).

Recherches sur l'image

- Présentation d'un projet concret de construction d'une interface informatique offrant un mode d'expression idéographique à des enfants paralysés mais disposant de l'ouïe et de la vue (p. 251).
- Problèmes de l'accès à une banque de données d'images avec l'appui de l'hypertexte (p. 379).
- Modèle d'analyse fine de photographies en vue de la classification d'une banque iconographique (p. 275).

Jacques Maniez et Widad Mustafa el Hadi