

ÎLE ROUSSE

12 mai au 16 mai 1997

Les systèmes d'information élaborée

*Bibliométrie - Linguistique - Information Stratégique - Veille Technologique
Intelligence économique*



*

PROGRAMME

*

*Journées d'études organisées par
la Société Française de Bibliométrie Appliquée
sous l'égide du
Ministère de l'Enseignement Supérieur et de la Recherche (DIST)*

Avec le soutien de :

*la Municipalité d'Ile Rousse
l'Université d'Aix-Marseille (CRRM)
l'Université de Marne-la-Vallée
le Centre de Documentation de l'Armement (CEDOCAR)
le Centre d'Études Scientifiques de Défense (CESD)*

COMITE SCIENTIFIQUE DU PROGRAMME

<i>H. DOU (CRRM)</i>	<i>P. LACOSTE (CESD)</i>	<i>F. MARCOTORCHINO (IBM)</i>
<i>B. DOUSSET (IRIT)</i>	<i>D. LAURENT (U.M.L.V.)</i>	<i>C. PAOLI (CESD)</i>
<i>J. DUCLOY (CRIN-Nancy)</i>	<i>Y. LE COADIC (CNAM)</i>	<i>L. QUONIAM (CRRM)</i>
<i>J.E. DUBOIS (Paris VII)</i>	<i>F. LEPINE (FED)</i>	<i>R. SOLER (CERE-CERAM)</i>
<i>C. DUTHEUIL (Synthelabo)</i>	<i>C. LONGEVIALLE (DGA)</i>	<i>W. TURNER (CERESI)</i>
<i>J. GUYAUX</i>	<i>B. MARX (INPI)</i>	<i>E. VALENSI (CEDOCAR)</i>

MERCREDI 14 mai 1997

- 9h00 Accueil des participants
- 9h15 Ouverture des journées :- P. PASQUINI, Ministre, Député - Maire d'Ile Rousse,
- P. LACOSTE, Président du C.E.S.D.,
- H. DOU, Président de la S.F.B.A.
- 10h00 Interventions inaugurales:- J-E. DUBOIS, Président de CODATA INTERNATIONAL,
- E. VALENSI, Directeur du CEDOCAR.

11h00 *Pause*

SESSION I: Stratégie et contexte

Président : H. DOU

- 11h15 **La classification neuronale pour la détection de nouvelles tendances de recherche et le développement de nouveaux produits.**
P. Bereau et H. Dou (CIFRE ANRT et CRRM) p 9
- 11h45 **Competitive intelligence for brazilian small and medium - sized companies.**
L. R. Fernandes, C. H. da Silva, G. M. Coelho, L. Quoniam et H. Dou
(Instituto Nacional de Tecnologia - Brazil et CRRM) p 10
- 12h15 **Évaluation de l'offre de nouveaux produits comme mesure des résultats de l'innovation.**
J.M. Belot et R. Dalpe (CETIM et CIRST) p 11
- 12h45 *Déjeuner*

SESSION II: Stratégie et contexte (suite)

Président: J. GUYAUX

- 14h30 **Brevets mondiaux déposés au Brésil sur le marché des technologies globales.**
N. Macedo dos santos et Milton de Abreu campanario
(Instituto de Pesquisas Tecnologicas do Estado de Sao Paulo) p 12
- 15h00 **Thomson Multimédia, un portefeuille de brevets plus que symbolique.**
D. Dore (Thomson) p 13
- 15h30 **Une mesure de l'intelligence.**
E. Valensi (CEDOCAR) nc
- 16h00 *Pause*

SESSION III: Stratégie et contexte (suite)

Président : Y. LECOADIC

- 16h15 **Réflexion sur l'évolution de la profession de veilleur.**
H. Ziegelbaum et P. Lauri (Tepral) p 14
- 16h45 **Competitive technical intelligence : human resources training and technological assistance to brazilian companies.**
G. M. Coelho, L. R. Fernandes, C. H. da Silva, V. L. M. Lellis, L. Quoniam et H. Dou
(Instituto Nacional de Tecnologia - Brazil et CRRM) p 15
- 17h15 **Aérospatiale : Mise en réseau de la veille technologique.**
M. Meyer (Aérospatiale centre commun de recherche Louis Blériot) p 16
- 17h45 **La mise en place d'un service de veille technologique à la norme ISO 9000.**
P. Lauri et Dou (Gemplus et CRRM) p 18
- 18h15 *Fin de session*

MERCREDI 14 mai 1997

(Session parallèle)

SESSION IV: Systèmes d'information

Président: L. QUONIAM

- 15h00 **Un lien effectif entre informations scientifiques et informations techniques.**
P. Faucompré, L. Grivel, X. Polanco, H. Dou et L. Quoniam
(CRRM université d'Aix-Marseille III et INIST) p 19
- 15h30 **Les processus informationnels dans une activité professionnelle.**
R. Palermi et C. Moriset
(IUT Grenoble II) p 29
- 16h00 *Pause*

SESSION V: Systèmes d'information

Président: L. PEREIRA

- 16h15 **Management des systèmes d'information et génie organisationnel.**
S. Dumas et P. Dumas
(Gister Innovation et IUT de Toulon et du Var) p 31
- 16h45 **La chimie en région PACA à travers les derniers congrès de la Société Française de Chimie SPC de 1994 à 1996.**
J. Kister, L. Quoniam et H. Dou (GOAE et CRRM) p 32
- 17h15 *Fin de session*

JEUDI 15 mai 1997

SESSION VI:

Méthodes et outils

Président: C. DUTHEUIL

- 9h00 **Vers une production automatisée de l'information élaborée : « l'intégration de robots dans la plate-forme ATLAS ».**
P. Andrei, T. Regnier, S. Aubert et C. Paoli (Aérospatiale, CEDOCAR et CESD) p 33
- 9h30 **Éléments méthodologiques par la conception de systèmes d'intelligence économique à l'entreprise considérés comme des systèmes d'aide à la décision.**
M. Salles, A.M. Alquier (INPT, département GSI) p 34
- 10h00 **Conception d'un nouvel outil de gestion des ressources hétérogènes en veille technologique.**
V. Léveillé, H. Rostaing et P. Hassanaly. (CRRM et Laboratoire de valorisation de l'Information) p 35
- 10h30 *Pause*

SESSION VII:

Méthodes et outils (suite)

Président : J. KISTER

- 10h45 **A software tool for bibliometric technology monitoring.** A. Kopcsa (Seibersdorf, Autriche)
- 11h15 **Liens entre distribution d'usage et de structure des documents.**
T. Lafouge et S. Laine-Cruzel (Laboratoire RECODOC) p 36
- 11h45 **Profil-doc : un système personnalisé de requête à des bases de données en texte intégral.**
C. Michel et T. Lafouge (Laboratoire RECODOC) p 37
- 12h15 **Homogénéisation et traitement de données multi-sources.** M-H Antoni (ECAM IBM-France) nc
- 12h45 *Déjeuner*

SESSION VIII:

Méthodes et outils (suite)

Président : F. MARCOTORCHINO

- 14h30 **Une méthode de traitement automatique des questions ouvertes dans une enquête consommateur.** H. Ziegelbaum, M. Rogeaux et H. Rostaing (TREPAL et CRRM) p38
- 15h00 **Les approches Multiparticipant Multicritère d'aide à la décision.**
R. Ginting et H. Dou (Institute of Technology of Indonesie et CRRM) p 39
- 15h30 **Maintenance d'une base de données techniques de référence : l'apport du veilleur à la fourniture d'information validée aux experts de l'entreprise.** D. Graveleau (DGA et CTSN)
- 16h00 *Pause*

SESSION IX:

Méthodes et outils (suite)

Président : C. PAOLI

- 16h15 **Méthodes globales d'analyse bibliométrique et statistique.** J. Guyaux nc
- 16h45 **Une approche pour la modélisation de la sémantique dans les systèmes d'information élaborée.**
P. Aniorte (Institut Universitaire de Technologie de Bayonne) p 40
- 17h15 **Flash-Poster : Utilisation des techniques bibliométriques dans l'analyse d'un dossier individuel de chercheur au CNRS : Limites et perspectives.**
J. Kister, L. Quoniam (GOAE et CRRM) p 41
- 17h30 *Fin de session*

JEUDI 15 mai 1997

(Session parallèle)

SESSION X: Internet

Président: E. VALENSI

- 9h00 **Recherche précise d'informations sur le www et veille technologique : utilisation de données structurées pour l'interrogation via les moteurs de recherche.**
J. Ducloy, E. Nauer et J-C Lamirel (CRIN, CNRS et INRA - Lorraine) p 43
- 9h30 **Internet : une source indispensable dans la réalisation des dossiers de veille.**
S. Dumas et J.M. Dou (CRRM et Gister Innovation) p 45
- 10h00 **Audit d'un serveur Internet et approche réseau.**
E. Boutin, H. Rostaing et L. Quoniam (CRRM et Laboratoire Lepont) p 47
- 10h30 *Pause*

SESSION XI: Internet

Président: P. DUMAS

- 10h45 **La veille et Intranet.**
C. Baffoux, J.- M. Dou et S. von Bandel (ECIA et CRRM) p 48
- 11h15 **Veille scientifique et technique sur Internet.**
B. Dousset, T. Dkaki et J. Mothe
(IRIT et IUFM Université Paul Sabatier de Toulouse) p 49
- 11h45 **AURESYS 2.0 : Un agent intelligent au service de l'information stratégique.**
B. Mannina, L. Quoniam et H. Dou (CRRM) p 50
- 12h15 **D'un outil d'assistance à la consultation de l'annuaire REDOC Grenoble vers un serveur d'indexation et de recherche sur le w.w.w.**
G. Eymard et J-M Francony (IUT Grenoble II) p 51
- 12h45 *Déjeuner*

SESSION XII: DEMONSTRATIONS

- 14h30 Démonstrations de systèmes et de logiciels

VENDREDI 16 mai 1997

SESSION XIII:

Applications

Président: B. DOUSSET

- 9h00 **Une approche bibliométrique des sciences sociales et humaines au Portugal.**
R. Couto Costa (Ministère des sciences et des technologies) p 53
- 9h30 **Évaluation et expertise scientifique.**
B. Dousset et T. Dkaki (IRIT Université Paul Sabatier de Toulouse) p 54
- 10h00 **Résistance aux maladies chez les plantes transgéniques.**
G. Branca-Lacombe, F. Cotte et A. Texier-Drouhard (INRA) p 55
- 10h30 *Pause*

SESSION XIV:

Applications

Président: H. DOU

- 10h45 **Analyse automatique de l'information scientifique pour créer une cartographie des interactions géniques.**
V. Pillet, L. Quoniam et B. Jacq
(CRRM et LGPD Université de Luminy, marseille) p 56
- 11h15 **Veille scientifique, veille technologique : application à l'économie et à la recherche en Tunisie.**
S. Mahmoud, F. Chichti et P. Hassanaly p 57
(Institut Supérieur de Documentation et Laboratoire de Valorisation de l'Information)
- 11h45 **Mise en place d'un observatoire thématique sur les bitumes**
J. Kister, L. Quoniam (GOAE et CRRM) p 58
- 12h15 *Déjeuner*

SESSION XV:

Applications (suite)

Président: W. TURNER

- 14h00 **L'adaptation de la veille technologique aux pays africains.**
O. Ouattara et H. Dou (CRRM) p 60
- 15h00 **Les sources d'information légales au Luxembourg.**
L. Pohl, S. Quazzotti (Centre de Recherche Public Henri Tudor) p 61
- 15h30 **Le marché des nouveaux médias en France.**
C. Massa, L. Cornu (Centre de Recherche Public Henri Tudor) p 62
- 16h00 *Pause*
- 16h15 **La R&D au Portugal**
Rodrigues Pereira Fernando Luis (Universidade Fernando Pessoa) p 63
- 16h45 **Vers de nouvelles formes d'informations élaborées.**
F. Jakobiak (Existrat) p64
- 17h15 **L'intégration du format P.D.F. dans l'élaboration des systèmes d'information.**
C. Dou et E. Giraud (CRRM et Conseil Régional PACA) p66
- 17h45 *Fin de session*

VENDREDI 16 mai 1997

(Session parallèle)

SESSION XVI:

Analyse des textes

Président: J. DUCLOY

- | | | |
|-------|---|------|
| 10h00 | Text mining application.
C. Huot (IBM-France) | p 67 |
| 10h30 | Impact scientifique des laboratoires d'un département du CNRS.
B. Ring et G. Vergnes (CNRS) | p 68 |
| 11h00 | L'accès à la presse scientifique médicale : mutation en cours.
G. Chartron (URFIST de paris et Ecole Nationale des Chartes) | p 70 |
| 11h30 | <i>Pause</i> | |
| 11h45 | Cartographie d'un champ de recherche à partir du corpus des thèses de doctorat soutenues pendant 20 ans : les sciences de l'information et de la communication en France 1974/1994.
Y. Polity et H. Rostaing (CRRM et IUT de Grenoble II) | p 71 |
| 12h15 | Coverage of national journals makes all the difference.
I. Gomez, M. Teresa Fernandez et F. Morillo (CINDOC et CSIC) | p 72 |
| 12h45 | <i>Déjeuner</i> | |

SESSION XVII:

Analyse des textes (suite)

Président: C. LONGEVIALLE

- | | | |
|-------|---|------|
| 14h30 | Optical switching : state-of-art.
F.H. Papa, C.A. Loural et FLC Miquelino (R&D labs of Brazilian Telecom) | p 73 |
| 15h00 | Gain de qualité sur l'information pour les entreprises avec l'utilisation en tandem des bases de données et d'Internet. Analyse et comparaison de 11 sujets scientifiques.
J.P. Brachet (IRIAM) | p 75 |
| 15h30 | Le journal télévisé de 20 heures (TF1) carrefour social.
B. d'Aiguillon (Université de la Méditerranée, Marseille) | p 76 |
| 16h00 | <i>Fin de session</i> | |
| 16h30 | <i>Fin des journées</i> | |

Competitive intelligence for brazilian small and medium- sized compaignies

L.R. FRENANDES, C.H. DA SILVA, G.M. COELHO, L.QUONIAM, H.DOUGLAS
Instituto Nacional de Tecnologia, C.R.R.M.

The strategy of the National Institute of Technology- INT lays on integrated actions towards the generation and dissemination of low costs and high aggregated value technologies, in order to attend social priorities and to provide technological support to micro, small and medium-sized compaignies.

In this context, the Technological Information Division of INT has been investing hardly in training its staff on new information technologies eand methodologies, mainly technological awareness, aiming the development of human resources devored to specialized technological information services, in order to attend its own researchers needs and the demand of Brazilian compaignies.

This paper shows an example of multi-clients study on technological awarness. The conclusions are based on bibliometric analysis.

The methodology presented could be adapted to increase the activities developed by small and medium-sized enterprises - SME's of any industrial sector in Brazil.

The survey of compaignies needs should be made through questionnaires and interviews with the technicians of each compaigny. The evolution of the study, based upon the obtained data, supported by statistical methods can lead to integrating the monitoring to the strategic aims of each enterprise.

COMPETITIVE TECHNICAL INTELLIGENCE : HUMAN RESOURCES TRAINING AND TECHNOLOGICAL ASSISTANCE TO BRAZILIAN COMPANIES

L.R. Fernandes, C. H. da Silva, G. M. Coelho, L. Quoniam, H. Dou
Instituto Nacional de Tecnologia, CRRM

The experience in different countries shows that the search for excellency and better performance contribute to the growing usage of information systems to support decision making and guarantee the reduction of time face to external factors exigency.

The role of competitive technical intelligence is to offer a systematic gathering and analysis of public information about the companies competitors, allowing its usage in decision making.

The CTI works as a radar, identifying new opportunities and avoiding disasters. CTI analysis identify signs from external environment and analyse them in relation to the companies needs and competencies, allowing forecast and technological, tactical and strategic planning.

This paper presents an outlook of CTI and the activities in development at the National Institute of Technology in Brazil, in cooperation with CRRM / Université d'Aix-Marseille - France, aiming to train human resources and to implement CTI systems in Brazilian companies.

Un lien effectif entre informations scientifiques et informations techniques

P. FAUCOMPRE, L.GRIVEL, X. POLANCO, H. DOU, L. QUONIAM
C.R.R.M., INIST

Si le lien entre science et technologie représente un enjeu stratégique majeur, la mise en relation de références bibliographiques scientifiques et de références bibliographiques techniques peut présenter un intérêt documentaire très important. Pour réaliser ce lien, les mots clés de la Classification Internationale des Brevets ont été utilisés comme pivot relationnel entre les vocabulaires d'indexation de grandes banques de données et les codes de classement des brevets. Une première étude avait montré la faisabilité d'une telle mise en correspondance entièrement automatique et ses insuffisances. Nous présentons ici quelques modifications importantes que nous avons apportées à cette mise en relation, notamment la prise en compte d'index multilingues permettant de lier plusieurs champs d'indexation à une représentation la plus complète possible de la classification des brevets.

L'évolution essentielle de notre projet concerne le mécanisme de mise en correspondance qui aboutit aujourd'hui à une réindexation globale d'une notice bibliographique à l'aide de codes de classement. Cette évolution concerne également le concept même de mise en correspondance qui doit être interprété comme une simple présomption de lien entre informations scientifiques et informations techniques. Ces développements ont quelques conséquences sur notre interface et son évaluation. D'abord, dans la mesure où il ne s'agit plus d'un lien biunivoque terme à terme, l'utilisateur n'a plus à sélectionner des mots clés intermédiaires générateurs de concordances mais peut directement exploiter les codes symbolisant la classification de la propriété industrielle. Ensuite, les principaux indicateurs documentaires semblent mal adaptés pour mesurer la performance de ce nouveau champ dont l'unique rôle est de proposer des pistes à explorer. Enfin, l'utilisateur final paraît être le seul acteur en mesure d'apporter une validation complète, et non seulement documentaire, de cette mise en relation.

Introduction

Les entreprises soucieuses d'innovations technologiques sont confrontées à la difficulté d'articuler rapidement une recherche d'informations techniques à une recherche d'informations scientifiques. Pour certains acteurs économiques, la représentation documentaire de ces deux univers et la multiplication des systèmes d'information pour y accéder affaiblissent la visibilité des liens existants ou recherchés entre la recherche scientifique et la recherche industrielle. Le passage de l'une à l'autre permet aux industriels de détecter plus rapidement les opportunités potentielles induites par l'activité scientifique et aux laboratoires de percevoir les besoins de solutions techniques émanant des entreprises. Pour accélérer ce passage, nous avons construit un système de mise en correspondance automatique entre ces deux domaines aboutissant à la projection de la Classification Internationale des Brevets sur le signalement de la littérature scientifique : chaque référence bibliographique réindexée par ce système comporte un nouveau champ contenant un ou plusieurs codes de classement spécifiques à la propriété industrielle. Dans une précédente communication (), nous avons montré les principes de faisabilité de ce

type de mise en relation. Après le rappel du contexte de ce projet, nous voudrions aujourd'hui, en exposer les premiers résultats ainsi que ses limites actuelles, en insistant particulièrement sur l'évolution du concept même de mise en correspondance.

L'enjeu d'une relation science et technologie

Les informations scientifiques et les informations techniques sont des composantes essentielles du transfert entre la recherche scientifique et le développement technique. Leur lisibilité contribue à accélérer leur échange (), notamment à travers les réseaux de collaboration où s'élaborent, en partie, les processus d'innovation technologique. Les entreprises participant à ces réseaux sont particulièrement sensibles aux liens réels ou souhaités entre informations scientifiques et informations industrielles. Mais la recherche industrielle redéfinit constamment les frontières établies entre la recherche académique et l'état de la technique. Si on la considère comme l'un des facteurs décisifs de la compétitivité et de l'avantage concurrentiel, la prise en compte la plus rapide possible des informations nécessaires à l'innovation devient un facteur déterminant dans la prise des décisions d'ordre stratégique de l'entreprise. Avec l'émergence de ces réseaux, disparaît une vision purement mécaniste et linéaire des flux existant entre la recherche, le développement et la production : chaque plan évolue selon une logique qui lui est propre (&). Dans ces élaborations communes () et ces configurations sans cesse remodelées, tous les acteurs nouent et dénouent des liens avec les autres, contribuant au rapprochement ou à l'éloignement des différents plans entre eux (). Chaque intervenant doit alors y appréhender un environnement toujours plus large et plus mouvant et pour maîtriser les relations qu'il est amené à tisser avec un ensemble de partenaires techniques ou économiques, il doit capter et analyser des informations très diversifiées, sonder de multiples sources et avoir recours à des outils et des méthodes de recherche extrêmement variées. C'est dans la perception de cette complexité que certaines entreprises modifient leur approche du développement technique () et leur attitude face à la recherche (). Car si elles parviennent à s'appropriier () la technologie sous forme de connaissances, de compétences ou de biens, elles ne le peuvent que dans la limite de leur culture technique (). Pour ces entreprises, la lisibilité des informations nécessaires à ces constructions collectives est une composante essentielle de toute collaboration. Sans cette lisibilité, elles risquent de se détourner de l'information technique et de ses procédures d'accès vécues comme « *barrières objectives* » (), ou de ne pas pouvoir identifier les solutions techniques et le potentiel innovateur que le développement scientifique pourrait leur apporter.

Il est certes reconnu que l'essentiel des rapports de collaboration tissés entre entreprises, ou entre industriels et laboratoires de recherche, s'établit préférentiellement au niveau de contacts informels à travers des réseaux personnels extrêmement proches. Ces relations apportent incontestablement un niveau élevé de crédibilité, de confidentialité et de confiance indispensable aux projets dans lesquels l'entreprise investit une part de ses ressources et parfois son avenir. Mais au delà de cette assurance, les échanges formalisés contribuent eux aussi à réduire une part de l'incertitude inhérente à ces collaborations. Les informations documentaires, parmi toutes les connaissances impliquées dans une démarche d'innovation technologique, jouent, d'une manière générale, un rôle important dans ces échanges et réorientent, en particulier, le travail du spécialiste de l'information dont la mission aujourd'hui n'est plus « *de maîtriser l'information mais de l'ouvrir, et cela de façon collective* » ().

La nécessité d'une passerelle automatique

En raison de l'abondance des systèmes documentaires et de leurs outils de recherche, notre projet s'appuie sur deux systèmes que l'on pourrait qualifier de canonique dans leur domaine respectif, la littérature technique, d'une part, et l'information scientifique, d'autre part.

• la représentation de l'information technique

Une passerelle entre science et technologie n'est pas réalisable par un lien simple et direct. Aussi est-elle bâtie, très indirectement, sur une mise en relation de références bibliographiques issues de la littérature scientifique et de références issues de la littérature technique, les documents de brevets, ceci par le biais de la Classification Internationale des Brevets (CIB). Rappelons que cette classification est employée par les examinateurs des Offices de propriété industrielle pour décider si la chose présentée comme nouveauté tombe sous le concept d'invention technique et mérite ainsi un titre de propriété industrielle. Son objectif principal étant de permettre d'établir des comparaisons entre des objets techniques, elle procède essentiellement au classement de l'objet technique dans le domaine auquel il appartient (&). La CIB est la représentation documentaire exhaustive du système de protection juridique des objets techniques légalement appliqué par 75 Offices de brevets nationaux et régionaux. De ce fait, plus de 20 millions de brevets accessibles dans le monde sont classés à l'aide de ces codes. L'intérêt de cette classification, couramment utilisée par les industriels, est d'autant plus grand que 70% à 80% de l'information technique se trouvent exclusivement dans les documents de brevet (&) et que ces documents, tous obligatoirement publics, sont signalés par près de 90 banques de données. Rappelons également, que sur le plan strictement documentaire, seule une classification monohiérarchique et énumérative comme l'est la CIB est capable, de nos jours, d'assurer l'exhaustivité d'une recherche.

• la représentation de l'information scientifique

Au niveau de l'information scientifique, la base de données PASCAL de l'INIST présente de nombreux avantages. Ses 11 millions de références bibliographiques, dont près de 85% issues de la littérature fondamentale, couvrent tous les grands domaines scientifiques et leur périphérie ainsi que de nombreux secteurs connexes. Son positionnement comme base de première approche, résolument multidisciplinaire, multilingue et internationale, lui donne une dimension exploratoire qui en fait une base expérimentale idéale pour notre projet. Sur le plan documentaire, les objectifs de cette base lui imposent naturellement un langage d'indexation très analytique caractérisé par un vocabulaire extrêmement volumineux et faiblement structuré.

• un langage pivot

On peut s'étonner du fait que les deux langages documentaires sélectionnés sont manifestement sinon totalement opposés tout au moins fortement hétérogènes dans leur nature, leur fonctionnement et leur objectif. Il semble évident que des outils reposant sur des principes si éloignés rendent impossible la construction d'un lien direct entre eux. La base PASCAL possède son propre plan de classement mais il est reconnu qu'il est très difficile, voire impossible (), d'établir une correspondance satisfaisante entre deux classifications reposant sur des principes de division logique très différents. Un langage intermédiaire compatible avec ces deux représentations est donc nécessaire. Qu'existe-t-il alors du côté de la CIB? Cette classification comprend des index de mots clés appelés catchwords. Les catchwords offrent à l'utilisateur un accès direct aux codes de classement, sans présupposer aucune connaissance préalable de la structure hiérarchique et apporte à l'examineur de brevets une aide éventuelle pour délimiter un sujet technique dans l'étape de préparation d'une recherche documentaire. Le

choix des catchwords s'impose alors pour trois raisons immédiates : en premier lieu, ils rendent inutiles la construction d'un nouvel outil, en second lieu, ils offrent une passerelle naturelle et validée de termes contrôlés avec les mots clés des autres sources documentaires, et enfin, en troisième lieu, ils permettent de s'affranchir d'un lien direct avec le texte de la classification proprement dite.

- **un lien automatique**

Les outils étant sélectionnés, il faut maintenant trouver un mécanisme pour corréler les éléments de ces deux systèmes et pour permettre, à partir de termes d'indexation ou de classification, de passer de l'un à l'autre. Aussi «*the following suggestions are intended to provide methods for setting up of a compatibility instrument, a black box*» (). Un traitement manuel ne peut pas être envisagé pour au moins 6 raisons d'ordre quantitatif et qualitatif :

1. l'importance quantitative des volumes à traiter
2. le coût financier exorbitant d'une telle opération (plusieurs dizaines d'année-hommes)
3. la formation aux doubles compétences en classification et en indexation
4. la difficulté à maîtriser les dérives liées à la durée d'une telle réalisation
5. le coût d'entretien et de mise à jour d'un tel système
6. la nécessité de rendre répétable ce traitement

Compte tenu de ces contraintes, le principe d'une telle mise en correspondance ne peut retenir que des options relativement basses : le lien est donc bâti uniquement sur la coïncidence lexicale entre les définitions complètes des catchwords et les descripteurs unitermes ou multitermes de la base cible. Il est important de souligner que cette simplicité permet d'apporter une relative robustesse à la mise en correspondance et assure un certain réalisme aux liens établis entre les deux systèmes de représentation documentaire. Ajoutons également que pour renforcer la légitimité de ce lien exclusivement construit sur la stricte identité des mots, le traitement de mise en correspondance, notre *boîte noire*, fait appel à une succession de filtres faisant notamment intervenir la considération de valeurs numériques donnant la mesure de certains paramètres comme le degré d'inclusion. Ces filtres jouent un double rôle : sur un plan qualitatif, ils permettent de sélectionner les expressions les plus proches au niveau lexical quand des situations de choix concurrents se présentent, ce qui est le cas de figure le plus fréquent, et, sur un plan apparemment strictement quantitatif, ils permettent d'éviter l'explosion du nombre de liens.

L'évolution du concept de mise en correspondance

Les possibilités et les insuffisances mises en évidence au cours de l'étude de faisabilité ont fait progressivement évoluer le projet de mise en correspondance, plus au niveau du concept lui-même qu'au niveau de sa technique. D'une manière apparemment paradoxale, cette évolution est marquée à la fois par la prise en compte de nouvelles données documentaires et par un certain éloignement de ce seul plan documentaire compris au sens strict du terme.

- **l'élargissement du vocabulaire technique**

Pour couvrir le plus grand nombre de domaines possible, et conforme à l'idée que «*The greater the relative specificity and vocabulary size of the input language, the better the switching performance*» (), le système de mise en relation n'utilise plus seulement les catchwords français mais les vocabulaires de 5 Offices de brevets nationaux (Allemagne, Autriche, France, Grande-Bretagne et Suisse) et conforte la perspective très européenne de cette classification. L'extension du nombre d'index permet de disposer d'un vocabulaire donnant une description la

plus complète possible de l'état de la technique et de respecter la diversité des gisements documentaires susceptibles d'une telle correspondance. L'élément le plus important n'est pas cependant le nombre absolu de définitions disponibles mais la représentation que ces 160 000 entrées offrent de la classification. Comparés aux 65 000 codes de la CIB dans sa totalité, les 3 index couvrent près des deux-tiers des objets et des domaines techniques classés. Si l'on considère un des niveaux hiérarchiques particulièrement important pour les recherches initiales, les 613 sous-classes présentes couvrent 99.51 % des codes. Le poids de l'index allemand, pris ici dans sa traduction anglaise, est bien sûr déterminant.

Origine de l'index	Langue	Définitions	Codes Couverture
France	français	19 881 14 328	22 %
Grande-Bretagne	anglais	18 029 9 663	15 %
Allemagne, Suisse, Autriche	anglais	122 81138 402	59 %
Σ	160 721	40 456	62 %

• **la prise en compte partielle de l'indexation**

Pour respecter au mieux l'esprit de l'indexation effective des références bibliographiques, la même démarche a été adoptée pour la base PASCAL. Les 3 types de vocabulaire utilisés, représentant l'indexation réelle de 3 années d'entrées dans la base, ont d'abord été considérés dans leur totalité. Mais leur analyse révéla que l'indexation libre présentait une caractéristique peu compatible avec une mise en correspondance : dans une distribution de type Zipf très prononcée, seuls 12 % des mots libres sont présents au moins une fois par année d'extraction.

Voculaires extraits	Champ*	Entrées	Doublons	Seuil	Traités
Descripteurs contrôlés	FD				
ED	96 433				
	99 046	1 388			
	3 153	-			
	95 045				
	95 893				
Table d'équivalence	F/E	128 717	3 508 -		125 210
Candidats descripteurs	FD				
ED	27 404				
	29 155	315			
	563	-	27 087		
	28 590				
Mots libres	FD				
ED	537 929				
	7 041	-			
	35	> 2			
	> 2	65 508			
	477				

* FD = French descriptor, ED = English descriptor

En d'autres termes, 88 % des mots libres indexent moins de 3 notices sur 3 années consécutives ce qui induit une espérance de lien quasi nulle. En conséquence, seules les entrées ayant une fréquence annuelle d'au moins 1 ont été retenues. Ce seuil présente également l'avantage de réduire d'autant le volume considérable de termes à traiter.

- **la reconnaissance de l'aspect multilingue des index**

La prise en compte des index disponibles en plusieurs langues découle du même élargissement de l'assise des vocabulaires. La justification du multilinguisme est cependant différente dans le cas de l'indexation et dans le cas de la CIB. Dans celui de la base PASCAL, la présence d'un terme dans les champs d'indexation français et anglais augmente le nombre potentiel de liens, notamment avec l'index allemand, et peut écarter certaines anomalies. Dans le cas des catchwords, la présence simultanée d'une entrée dans deux ou trois index multinationaux renforce de manière significative la probabilité de sa pertinence. En cas de concurrence de plusieurs entrées prétendant à un lien, celles ayant la présence la plus fréquente dans plusieurs index sont sélectionnées. Dans ce cas de figure, très majoritaire, le lien final correspond à un objet multi-index.

- **la suppression d'une interface visible**

Dans la version initiale élaborée pour l'étude de faisabilité, l'utilisateur était mis en présence de trois informations : les catchwords liés aux descripteurs propres à la référence bibliographique, et, associés à chacun de ces catchwords, un indicateur numérique indiquant un degré de pertinence ainsi qu'un ou plusieurs codes de classement. L'utilisateur devait ensuite sélectionner le catchword le plus intéressant, éventuellement en fonction du degré de correspondance proposé, et choisir dans un troisième temps le code de classement qu'il jugeait pertinent. Mais ce système était trop lourd en raison du nombre de niveaux intermédiaires et du nombre de sélections proposées. Le dispositif de mise en correspondance, en imposant une surcharge d'accès, au lieu de favoriser la lisibilité finale ne faisait qu'augmenter l'opacité globale des liens. Notre projet avait donc sous-estimé le fait que les utilisateurs seraient peu intéressés par des listes de mots clés supposés communs mais qu'aucun d'eux n'a coutume d'utiliser. L'outil relationnel doit donc être rendu entièrement transparent et seuls les codes de classement obtenus doivent être visibles. En faisant abstraction de l'outil de médiation, les utilisateurs peuvent directement interpréter les représentations documentaires qui leur sont familières sans se soucier des conditions de production des codes symbolisant la propriété industrielle. La notion, parfois suspectée, de boîte noire a toute sa raison d'être ici.

- **le concept d'une réindexation globale**

La précédente interface, d'une part, et l'utilisation d'un seul index et d'un seul champ d'indexation, d'autre part, obligeait chaque lien réalisé entre un descripteur et un catchword à être présenté isolément. Les codes de classement, uniquement accessibles à travers la sélection d'un catchword particulier, restaient totalement indépendants les uns par rapport aux autres. La nécessité, plus conceptuelle qu'ergonomique, de les présenter seuls a créé un rapprochement en direction du profil moyen des codes attribués aux documents de brevets. Dans la mesure où il n'existe pas de possibilité réaliste de privilégier un descripteur par rapport aux autres et étant donné qu'il n'y a pas de légitimité convaincante à calculer une valeur quelconque de mesure de distance ou de proximité entre deux descripteurs présents (), aucune synthèse ou classification automatique de l'indexation ne permet de sélectionner un ou plusieurs codes de classement. De plus, cette synthèse ou cette classification n'aurait pas permis d'exploiter le point commun entre tous ces codes qui est leur appartenance à une même hiérarchie. Le traitement associant les codes de classement à chaque référence tient compte maintenant de la fréquence hiérarchique des codes sélectionnés pour la réindexation finale, sans plus se préoccuper des catchwords intermédiaires devenus alors inutiles.

Mise en correspondance et réindexation

• le lien avec la base PASCAL

Le système générateur des correspondances étant opérationnel, les 3 vocabulaires cibles de la base PASCAL peuvent être maintenant automatiquement liés. Ce traitement est entièrement reproductible sur le vocabulaire d'indexation de n'importe quelle autre banque de données bibliographiques. Le système de correspondance a créé un total de 8 175 liens. Si ce nombre peut paraître faible au regard des volumes initiaux, il faut rappeler, d'une part, que chacun des deux vocabulaires est issu d'un univers totalement différent, et d'autre part que les outils documentaires conçus depuis une trentaine d'années, à l'exception de la CIB précisément, comptent en moyenne environ de 4 à 6 000 entrées. Au delà du score particulier de chacun des 3 types de vocabulaire, il importe de souligner leur contribution respective dans le nombre final de liens obtenus. En effet, après dédoublement, les trois sources totalisent 3 126 descripteurs français liés et 5 049 descripteurs anglais liés. Mais il est remarquable de constater que les descripteurs contrôlés fournissent à eux seuls près de 84 % des liens français et près de 96 % des liens anglais. Ce qui souligne la contribution extrêmement faible des candidats descripteurs et des mots libres alors que leur volume respectif correspond à 30 % et 68 % des descripteurs contrôlés. La prise en compte de ces deux vocabulaires, apparemment beaucoup trop spécifiques, pourrait être sans doute discutée. Précisons également que les 499 sous-classes présentes dans les codes couverts par les liens établis permettent de couvrir plus de 81 % de la hiérarchie de la CIB.

Vocabulaires cibles *Champ Entrées liées*

Descripteurs contrôlés FD

ED 2 619

4 843

Candidats descripteurs FD

ED 360

770

Mots libres FD

ED 1 589

16

• un exemple de réindexation

Les tables de correspondances ainsi créées ont été appliquées à la base de l'Observatoire de la Région Provence-Alpes-Côte-d'Azur. Cette base, en cours de constitution, compte actuellement près de 5 000 références bibliographiques. Elle a été générée par l'extraction de toutes les références de la base PASCAL de l'INIST satisfaisant uniquement au critère géographique du site de Marseille pour les années les plus récentes (1993-1995). L'objectif de cet observatoire est d'offrir aux entreprises, aux acteurs économiques et aux décideurs, publics ou privés, une image de la production des travaux scientifiques des laboratoires implantés localement afin de favoriser certaines opportunités de collaboration. Ces références bibliographiques ont donc été réindexées et deux-tiers d'entre elles contiennent aujourd'hui, à titre expérimental, des codes de classement CIB. La figure 1 montre les principaux résultats de cette réindexation. Pour mieux mesurer l'information nouvelle apportée par la correspondance, nous présentons, par ordre décroissant, quelques codes du plan de classement PASCAL et quelques descripteurs les plus fréquents présents dans les références. Au-dessous de ces deux séries, nous montrons les quelques codes CIB (5ème édition française) les plus fréquents accompagnés du titre principal très abrégé de leur définition.

Comme on peut l'observer sur cette figure, les codes CIB donnent une image relativement différente de celle offerte par les codes de classement PASCAL. Il est intéressant de noter que le système de correspondance ne semble pas introduire d'aberrations flagrantes par rapport à l'indexation attribuée par les indexeurs de l'INIST (ce qui paraît logique puisqu'elle provient des descripteurs). Cette nouvelle donnée ne se substitue à aucune information existante ni n'en corrige aucune. Tout au contraire, sa qualité dépend non seulement du traitement qui la produit mais également de tous les éléments de l'indexation. En tant que simple probabilité de lien, elle vient seulement compléter les informations présentes et enrichir l'interprétation que l'on peut donner d'une référence bibliographique en lui ajoutant une perspective d'ouverture vers d'autres domaines. Prenons un exemple : dans cette base, plusieurs publications traitent de pathologie rénale extrêmement aiguë. Les discussions y sont axées sur des problématiques d'actes opératoires, de symptomatologie ou des questions d'ordre méthodologique. L'indexation en rend parfaitement compte sur le terrain de la seule démarche scientifique, et c'est le rôle que l'on attend d'elle. Mais cette rigueur, nécessaire sur le plan fondamental, n'apporte pas une visibilité immédiate à certains utilisateurs non familiers avec ce type de vocabulaire, les PME/PMI, par exemple. Si l'on examine maintenant les codes CIB associés aux références, leur définition met en évidence le domaine couvrant les techniques de la dialyse et des reins artificiels. Cela ne signifie évidemment pas que la dialyse ou le rein artificiel soit la réponse à donner aux publications précédemment évoquées, mais qu'il existe une relation possible, que d'aucuns pourraient juger naïve, entre ces informations. Cette passerelle proposée entre des publications médicales, que l'on ne pourrait pas juridiquement traduire en termes de brevet, et le monde de la propriété industrielle est présentée sous une forme directement assimilable par les industriels. Sur le plan documentaire à proprement parler, il est important de remarquer que l'indexation native de ces références ne contient pas les descripteurs *reins artificiels* ou un concept équivalent. La figure 2 présente une notice complète avec son nouveau champ de codes CIB associés qui établissent un pont entre littérature scientifique et information technique. Pour être très précis, ces codes établissent seulement une relation avec le système de représentation documentaire de la protection juridique des inventions techniques. Mais dans la mesure où ce dernier couvre à lui seul 80% de l'information technique publique et publiée, on peut admettre qu'ils sont pratiquement équivalents.

Toutes les références contenant ce nouveau champ offrent l'opportunité de réinterroger les banques de données, spécialisées ou non dans les documents de brevet, contenant un champ de classement CIB à partir des codes qui leur ont été associés. L'exemple de la figure 3 illustre un lien qui peut être établi entre une publication scientifique référencée dans la base PASCAL et une référence de brevet signalée par les Chemical Abstracts en utilisant uniquement un code de classement. La première présente le travail d'un laboratoire d'océanologie de Marseille mettant en évidence l'influence de la concentration en sel marin sur le développement d'une bactérie dégradant les hydrocarbures et la seconde un brevet de la société Canon protégeant une nouvelle famille de bactéries capables elles aussi de dégrader certains composés aromatiques polluant l'eau ou les sols.

Discussions et perspectives

• un nouveau champ documentaire

Dans sa forme actuelle, le système de correspondance crée un nouveau champ contenant les codes de classement. D'une manière spécifique, l'utilisateur, sans même être très familiarisé avec la symbolique de cette classification technique, peut réemployer les codes obtenus dans des recherches sur des gisements de brevets. Inversement, un industriel pourra, à partir de ces seuls codes de classement et sans particulièrement maîtriser certains vocabulaires académiques,

remonter vers des publications d'origine plus fondamentale. D'une façon plus générale, ce champ peut être traité comme toute information disponible dans une référence et être combiné avec n'importe quel autre champ documentaire. Cependant, la limite évidente de ce lien produit automatiquement doit fixer également celle de toutes les interprétations que l'on peut en tirer, notamment dans son croisement avec d'autres informations et plus particulièrement dans son éventuelle exploitation bibliométrique.

- **le concept de simple présomption de lien**

Dans l'univers de la propriété industrielle, la notion même de concordance entre la CIB et les différentes classifications nationales est particulièrement problématique. Les conditions liées aux brevets ne sont pas d'un accès très aisé. «*They imply a degree of complexity which can in no way be covered by fine level of concordances*» (). L'évolution du concept de mise en correspondance dans notre projet est partiellement corollaire à l'abandon de l'idée de concordance en tant que telle. Cette correspondance ne doit donc pas être comprise comme un lien biunivoque, terme à terme, entre un mot clé et un autre mot clé, mais comme l'indice d'une relation possible entre des publications scientifiques et des brevets. Si le lien proposé n'est pas estimé aberrant et quand bien même il serait jugé particulièrement pertinent, l'interprétation ne peut pas franchir le pas de la présomption de lien à une certitude quelconque.

- **la pertinence des indicateurs d'évaluation documentaire**

En limitant la définition de cette relation à une simple présomption de lien, certains indicateurs, par ailleurs fort utiles et nécessaires dans l'évaluation des systèmes d'information documentaire, perdent une partie de leur pertinence. Au regard des indicateurs réputés être les plus utilisés, les plus facilement reproductibles et les moins controversés, cette relation ne peut plus être mesurée en termes de taux de rappel ou de taux de précision. A priori, seul l'indicateur de bruit conserve un sens. Cependant, cet indicateur ne peut guère indiquer le degré d'ouverture apporté par ce système, seule mesure utile de son degré d'intérêt et de son efficacité réelle.

- **validation multiexpertises et utilisateur final**

Comment également évaluer correctement ce système ? En amont, un strict contrôle documentaire des relations nouvellement proposées ne peut apporter aucune garantie de pertinence sinon seulement de constater que des mots communs ou des locutions communes ont été effectivement retrouvés dans les deux vocabulaires. Quant aux deux langages documentaires liés, chacun d'eux a déjà été amplement validé. En aval, chaque acteur intéressé par cette mise en correspondance excelle dans son domaine d'expertise mais ne peut, *à lui seul*, évaluer les liens établis. Une validation conjointe des multiples expertises doit alors s'établir dans un dialogue commun entre experts de domaines différents, experts qui ont à se livrer à un échange contradictoire. Cependant, la relation entre science et technologie est déjà quotidiennement réalisée par nombre d'utilisateurs finaux, et cette passerelle, toujours présente dans les processus d'innovation technologique, constitue leur propre compétence. En dernière instance, la place centrale revient donc à l'utilisateur final : son expérience est au coeur même du système qui lui est proposé et c'est précisément celle-ci qui peut donner une cohérence finale aux multiples expertises intermédiaires.

- **perspectives actuelles**

La troisième phase de ce projet devrait être l'étape d'une validation préindustrielle. Elle devrait porter sur 5 années de la base PASCAL, soit plus de 2,5 millions de références bibliographiques.

Conclusion

La relation entre information scientifique et information technique est intellectuellement et stratégiquement extrêmement complexe. Le réalisme d'une solution automatique, par nature fortement réductrice, impose un mécanisme de mise en correspondance qui soit simultanément très sélectif et uniquement basé sur des données de bas niveau. Il faut donc à la fois éviter l'explosion du nombre de liens et garantir une qualité statistique minimale des relations. La mise en oeuvre d'une série de filtres privilégiant les vocabulaires les plus fréquents et les codes sélectionnés hiérarchiquement les plus significatifs permet de lever cette apparente contradiction. Au niveau strictement statistique, le concept de simple présomption de lien non seulement s'éloigne nécessairement de la notion classique de pertinence mais également de la critique bibliométrique, par ailleurs puissante, des très fortes ou très faibles fréquences. Si la seule probabilité de lien génère un bruit documentaire évident, bruit qui doit être réduit, il n'existe cependant aucune raison statistiquement justifiable, et autre que le coût des traitements, de privilégier un palier de fréquences par rapport à un autre. L'interprétation des relations proposées et leur sélection repose alors entièrement sur le jugement de l'utilisateur : à lui de décider si celles-ci présentent un intérêt potentiel et si les voies suggérées méritent d'être explorées. Une telle mise en correspondance permet seulement à l'utilisateur d'explorer des pistes tracées entre deux univers conceptuellement bien distincts. En présence des autres champs documentaires couramment présents dans les références bibliographiques, les critères et les procédures de choix de l'utilisateur sont-ils si différents ?

LA CHIMIE EN REGION PACA A TRAVERS LES DERNIERS CONGRES DE LA SOCIETE FRANCAISE DE CHIMIE - SFC de 1994 à 1996

J. KISTER, L. QUONIAM et H. DOU
G.O.A.E., CRRM

Analyser la production scientifique est un moyen d'estimer les tendances de différents domaines, pays, journaux,... en vue de situer, suivre et adapter, voire programmer la politique de recherche d'un laboratoire, d'une Université ou d'un Organisme. L'information Scientifique et Technique (I.S.T.) apparaît ainsi de plus en plus comme un véritable outil de stratégie de recherche.

Les données bibliométriques permettent également d'identifier les équipes régionales, nationales ou internationales qui développent la même approche dans la même thématique ou dans un domaine proche. « Qui fait quoi, où, avec qui, comment et depuis quand, quelle évolution ? ».

Les deux derniers congrès de la SFC PACA de 1994 et 1995 vont donc servir de corpus de références pour analyser la chimie en région à travers une production scientifique particulière : les communications orales et posters. Les informations recueillies donnent une image des Unités de recherche CNRS ou laboratoires Universitaires, de leurs collaborations éventuelles. Les équipes de recherche et les liens entre celles-ci apparaissent et peuvent être suivies d'année en année mettant en évidence des collaborations ponctuelles ou établies ou des divisions issues de l'émergence de nouveaux groupes ou conséquences de restructurations ou d'incitations par les organismes de tutelle. Les congrès régionaux comme la « journée de la SFC PACA » traduisent relativement bien l'état de la chimie en région. Les contraintes d'acceptation des communications sont assez souples pour permettre à un grand nombre de jeunes étudiants et chercheurs de présenter l'activité des laboratoires à un temps t. Les observations réalisées sont comparées à celles issues des rapports de l'Université d'Aix-Marseille III (rapport d'UFR) pour la partie de la chimie sur le campus de St Jérôme. Dans ce cas, le corpus bibliographique est plus complet, il intègre les publications, les communications et conférences et des données plus structurelles comme le nombre de permanents, de thésards, les crédits récurrents etc...

Nous nous sommes particulièrement attachés pour ces deux corpus à la visualisation des réseaux d'auteurs et à leur évolution.

UTILISATION DES TECHNIQUES BIBLIOMETRIQUES DANS L'ANALYSE D'UN DOSSIER INDIVIDUEL DE CHERCHEUR CNRS : Limites et Perspectives.

J. KISTER, L. QUONIAM.
GOAE, CRRM

Le dossier scientifique des chercheurs se caractérise principalement par la production de publications, rapports techniques, thèses, proceedings, brevets et communication-conférences. Cette production restituée dans son contexte structurel (université, laboratoire, équipe, groupe de recherche), dans son contexte financier (moyens du chercheur ou de son laboratoire, contrats industriels ou institutionnels) et dans la tradition de son secteur scientifique (chimie, SHS, SPI, SDV...) peut-être analysée par les outils bibliométriques traditionnels.

Le dossier de production scientifique peut alors être considéré comme un corpus de références, élément caractéristique de la carrière du chercheur. Ce corpus est alors analysé quantitativement et qualitativement à travers les différents champs qui caractérisent cette base de données. La comparaison de la production papier et de l'aspect communication / conférences est un élément intéressant pour suivre et identifier les phases de la carrière du chercheur et son comportement comme chercheur de base ou comme moteur d'une équipe, éventuellement comme coordinateur de projets. La répartition dans le temps des différents supports de cette production permet aussi de déterminer le caractère de transfert de technologique en fonction du type de document ou du type de journal. Les journaux sont analysés dans leur diversité ou dans leur spécificité sans référence aux citations index. Il conviendrait à ce niveau d'avoir l'image cible du secteur, du domaine, ou de la thématique spécifique pour entrer dans une phase épistémologique du dossier.

L'analyse des réseaux d'auteurs, des fréquences moyennes ou des distributions du nombre des signataires est un élément d'analyse complémentaire surtout si l'on compare la production papier et la production communication/conférences. Cette présentation des réseaux d'auteurs peut se faire par thématique quand il y a une mobilité thématique. Il en est de même du type de document ou du type de journal et du champ langue.

Enfin la mobilité géographique ou thématique est aisément identifiée comme peut l'être une reconversion forcée ou programmée. Le changement de laboratoire peut-être visualisé quand il correspond à une volonté scientifique et non une stratégie purement opportuniste ou politique. De même les actions incitatives du CNRS ou du MESR peuvent être analysées.

Les limites de ce travail correspondent très vite aux limites des champs qui caractérisent cette production. Ainsi à l'interprétation, on retrouve les collaborations nationales ou internationales mais celles-ci ne peuvent être clairement identifiées par le manque d'affiliation des différents auteurs.

L'absence d'index-term ou de mots clés, limite l'analyse scientifique, celle-ci a été réalisée en posant le vrai problème de l'expertise, par rapport à un domaine scientifique.

En conclusion, ce travail sert de base de réflexion pour une présentation standardisée du dossier individuel d'un chercheur. L'utilisation de logiciels de base de données, mors de la constitution du dossier scientifique, devrait permettre d'aller plus loin et plus facilement dans l'analyse. La comparaison avec le corpus de références issu de l'interrogation de bases de données adaptées, devrait permettre de valider tout ou partie de la production papier. En ce qui concerne les communications - conférences, il convient de séparer les interventions régionales ou nationales, des congrès internationaux et des conférences sur invitations. Les évolutions comparées peuvent traduire une maturité du chercheur dans son groupe, son domaine ou sa thématique.

AUDIT D'UN SERVEUR INTERNET ET APPROCHE RESEAU.

E. BOUTIN, H. ROSTAING, L. QUONIAM

Laboratoire Lepont, C.R.R.M.

Grâce au WEB, une entreprise, une université, une organisation a la possibilité de mettre à la disposition de la communauté sur internet une information qui se présente sous forme de pages au format html qui s'enchaînent par un processus hypertexte. L'objectif recherché est suivant le cas de constituer une vitrine de ses activités, de répondre à une stratégie d'image ou de prestige, d'appuyer une stratégie commerciale ou d'être présent sur un nouveau médium. Un serveur ainsi constitué est consulté par des utilisateurs aussi appelés clients. Il est possible d'appliquer à cette confrontation entre l'offre et la demande d'information une des règles élémentaires du marketing: l'offre doit s'adapter à la demande car un site n'est pertinent que s'il est consulté.

Il est donc important pour l'organisation de disposer de capteurs dans son environnement qui lui permettent d'obtenir des informations statistiques correspondant à la visite de son site. Nous nous intéresserons ici à une source d'information incomplète mais toujours disponible. Elle résulte de l'utilisation du fichier Log qui enregistre les connexions des différents utilisateurs. L'analyse du fichier Log, sur une période donnée, peut permettre de dégager des invariants et doit fournir des réponses aux questions suivantes:

- Quelles sont les pages les plus visitées?
- Sur quelles pages les clients restent-ils le plus?
- Quel est le temps moyen passé par un utilisateur lors d'une connexion?
- A partir de quelle page les utilisateurs arrivent-ils sur notre site?
- A partir de quelle page les utilisateurs quittent-ils notre site?
- Quels sont les thèmes de notre site qui sont les plus consultés?

Ces éléments statistiques fournissent de précieuses indications sur la façon dont le site est utilisé et permettent au site de s'adapter aux besoins.

Il existe un certain nombre de logiciels commerciaux, connus sous le nom d'analyseurs de Log qui, partant du fichier Log proposent un ensemble d'indicateurs statistiques sur l'utilisation du serveur. Cette étude a pour objectif de renouveler l'approche traditionnelle des analyseurs de Log en utilisant l'analyse en terme de réseau. Cette nouvelle démarche est présentée en s'appuyant sur l'exemple de l'audit du serveur du Crrm réalisé en Décembre 1996 à partir de 2800 connexions.

Dans un premier temps, nous positionnerons la méthode réseau par rapport aux méthodes existantes.

Dans un second temps, nous développerons un bref aperçu de la richesse des analyses qui peuvent être conduites par cette méthode.

AURESYS 2.0

Un Agent Intelligent au Service de l'Information Stratégique

Mannina Bruno, Quoniam Luc, Dou Henri
CRRM

Internet est aujourd'hui une source d'information impossible à négliger: Cinquante millions d'ordinateurs y sont connectés, qui peuvent offrir des bases de données ainsi que des sources d'informations informelles.

Au regard de cet énorme flux d'informations, il est primordial d'y rechercher et collecter de l'information sans s'éloigner du sujet à traiter et tout en étant suffisamment exhaustif.

Devant la taille de ce réseau, certains outils permettant aux utilisateurs de circuler sur les Inforoutes, doivent être considérés de nos jours comme incontournables. Ces outils sont appelés **Agents Intelligents**, ou plus couramment Robots.

Les agents intelligents sont des programmes capables de réagir avec un environnement, de s'adapter aux circonstances, de prendre des décisions ou d'enrichir eux-mêmes leur propre comportement, sur la base d'observations qu'ils effectuent. Les progrès en la matière sont très rapide, cependant peu de robots présentent toutes les caractéristiques suffisantes pour des applications en **veille technologique** et/ou en **analyse automatique**. Certains moteurs, tel que Lycos, Yahoo, etc... ne sont généralement pas très adéquate pour ce travail car ils ne sont pas paramétrables avec la souplesse nécessaire.

Auresys est un agent intelligent développé au CRRM qui permet à ses utilisateurs de parcourir l'ensemble des Inforoutes en recherchant les informations conformes aux besoins stratégiques de l'utilisateur. Auresys, par rapport aux différents agents existant sur l'**Internet**, se positionne comme un des rares moteurs intelligents répondant aux préoccupations nécessaires pour l'activité. Nous présenterons un comparatif des différents agents existants sur les inforoutes et leurs différentes fonctions, avec le positionnement concurrentiel d'Auresys.

Nous montrerons que cette méthode de recherche d'information permet un gain de temps et d'argent.

Nous démontrerons qu'à partir d'une récolte d'information depuis l'Internet, il est possible de créer des bases de données personnalisées et consultables à distance. A partir de ces bases de données personnalisées, nous montrerons aussi qu'il est possible d'affiner son interrogation en local afin d'avoir une information plus ciblée. Nous présenterons des exemples de traitements automatiques de l'information à but infostratégique.

Nous finirons en articulant cette source d'information et son traitement dans un processus de veille en développant le caractère indispensable de cette source d'information.

ANALYSE AUTOMATIQUE DE L'INFORMATION SCIENTIFIQUE POUR CREER UNE CARTOGRAPHIE DES INTERACTIONS GENIQUES.

V. PILLET*, L. QUONIAM*, B. JACQ**

*CRRM, **Lab. de Génétique et Physiologie de Développement

Depuis la fin des années 1960, le flux d'information suit une croissance exponentielle. La complexité des informations et leur nombre sans cesse croissant ne permettent donc plus une lecture séquentielle classique du stock des données. Or il est nécessaire pour tout organisme de savoir maîtriser, organiser et gérer ce flux d'information. Il faut donc pour cela, se centrer sur l'utilisation de systèmes d'analyses automatiques des données afin d'extraire de cette masse celles qui sont à caractère stratégique.

Il existe de nombreuses sources d'informations, mais les concepteurs de bases de données conçoivent rarement leurs bases en fonction des applications futures qu'elles peuvent avoir. Tel est le cas dans le domaine de la génétique. Dans les bases de données sur les gènes et les protéines, le stockage et l'organisation des informations sont centrés sur les entités (gènes et protéines) considérés individuellement. Or il devient de plus en plus important de regarder les relations entre protéines, entre gènes et protéines pour découvrir leurs aspects dynamiques. La connaissance de ces interactions est indispensable pour comprendre les mécanismes mis en jeu lors du fonctionnement des systèmes biologiques.

Il a été mis en place une méthodologie qui permet d'extraire de façon automatique les interactions géniques existantes sur la *Drosophile* (mouche du vin), et ceci à partir d'une base de données générale sur cet insecte. Nous allons vous exposer les différentes étapes qui sont nécessaires à l'extraction de ces données et quels sont les résultats qui ont été obtenus. Cette méthodologie de traitements automatiques de l'information peut se subdiviser en trois parties. La première consiste à appliquer une série de traitements de reformatage des données pour créer une nouvelle base mieux structurée et débarrassée d'informations inutiles. La deuxième permet à partir de plusieurs critères de sélection, de ne retenir que les données les plus pertinentes. Enfin, un traitement sémantique a été appliqué sur ces données. L'analyse statistique de ces données a permis de mettre en évidence des groupes de mots spécifiques aux interactions géniques. Ces résultats ont permis d'affiner les critères de sélection nécessaires lors de la deuxième partie du traitement. Par ailleurs, grâce à un logiciel de traitement infographique nous avons pu établir une représentation graphique des interactions géniques repérées au niveau des données et qui en donne une vue synthétique.

Cette méthodologie de traitement et d'analyse de l'information a permis d'extraire d'une masse importante de données, des informations pertinentes à caractères stratégiques pour le chercheur et qui seront utiles pour la prise de décision. Cette méthodologie est donc un outil d'aide aux experts du domaine biologique. Elle permet de résumer et de synthétiser l'information sous forme de grille de lecture.

UTILISATION DES TECHNIQUES BIBLIOMETRIQUES DANS L'ANALYSE D'UN DOSSIER INDIVIDUEL DE CHERCHEUR CNRS : LIMITE ET PERSPECTIVE

Jacky KISTER, Luc QUONIAM
G.O.A.E., CRRM

Analyser la production scientifique est un moyen d'estimer les tendances des différents domaines, et ainsi d'orienter la politique de recherche d'un laboratoire ou d'un établissement.

Cette recherche, qu'elle soit fondamentale ou appliquée, doit de toute évidence se positionner autant dans le contexte national qu'europpéen ou international, en prenant en considération les publications scientifiques et les brevets. L'Information Scientifique et Technique (I.S.T.), ainsi que son traitement bibliométrique, apparaissent donc comme des outils indispensables à la prise de décision opérationnelle.

En effet, les analyses résultantes apportent une contribution non négligeable dans l'orientation de la politique recherche, quel que soit le domaine considéré. Elles sont particulièrement utiles pour lancer ou développer une technique ou pour montrer l'intérêt d'exploiter un thème précis par rapport à un autre.

Le chercheur du domaine considéré peut aussi utiliser ces outils pour adapter sa perception dans un domaine de spécialistes, hiérarchiser les supports d'informations leaders ou spécialisés. Il peut également rechercher l'aspect innovant dans des thèmes connexes.

Dans un domaine qui semble aussi pointu que la recherche sur les bitumes, l'approche Sciences de l'Information est intéressante, car elle visualise l'importance stratégique de ce domaine et sa grande diversité industrielle et scientifique.

La méthodologie adoptée comprend deux étapes, une analyse quantitative de la production scientifique dans le domaine par interrogation de bases de données adaptées et une analyse qualitative par traitement bibliométrique des informations téléchargées.

La première étape consiste donc dans l'élaboration d'un dossier de recherche documentaire, réalisé par association des experts du domaine et de ceux de l'IST. On définit ainsi un thème de recherche général et des sous-thèmes associés que l'on peut ou non croiser avec des concepts (vieillessement, biodégradation, rhéologie, radiolyse/photooxydation, technique d'analyse). On visualise ainsi, en terme d'évolution les grandes motivations de ce sujet, comme par exemple : les nouveaux revêtements routiers, les bitumes d'étanchéité, les enrobés de déchets ultimes ou sensibles.

Dans un deuxième temps le croisement, thème général ou sous-thèmes avec le concept vieillissement, fait l'objet d'une étude bibliométrique.

Ces analyses permettent de répondre aux questions posées par les experts du domaine, notamment en ce qui concerne l'identification des équipes régionales, nationales ou internationales qui se développent une même approche et avec quels partenaires industriels.

Cette étude n'est pas close puisqu'elle fait l'objet d'études complémentaires liées à des nouveaux développements contractuels issus de l'étude initiale.