



HAL
open science

Système intégré d'analyse automatique de l'information structurée

Henri Dou, Parina Hassanaly, Luc Quoniam, Hervé Rostaing, Albert La Tela,
Eric Giraud

► **To cite this version:**

Henri Dou, Parina Hassanaly, Luc Quoniam, Hervé Rostaing, Albert La Tela, et al.. Système intégré d'analyse automatique de l'information structurée. Veille Stratégique, Scientifique et Technologique, VSST'95, IRIT, Université de Toulouse, Oct 1995, Toulouse, France. pp.425-429. hal-01579953

HAL Id: hal-01579953

<https://hal.science/hal-01579953>

Submitted on 1 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SYSTEME INTEGRE D'ANALYSE AUTOMATIQUE DE L'INFORMATION STRUCTUREE

H. Dou, P. Hassanaly, L. Quoniam, H. Rostaing, A. La Tela, E. Giraud

*Centre de Recherche Rétrospective de Marseille (C.R.R.M.)
Université d'Aix - Marseille III - Centre Scientifique de St-Jérôme
Ave Escadrille Normandie - Niémen, 13397 Marseille cedex 20
Tél 91 28 87 11 - Fax 91 28 87 12
e-mail : crrm@crrm.univ-mrs.fr*

Résumé :

Produire des informations élaborées à partir de l'information structurée provenant de bases de données internes ou externes (scientifiques, brevets, technico-économiques, ...) nécessite la mise en place d'une chaîne de traitements faisant appel à différentes catégories de programmes informatiques.

Dans le cadre de cette approche, le CRRM a développé depuis plusieurs années différents outils, qui tout en s'appuyant sur un certain nombre de logiciels existants permettent un travail très précis, modulable, ouvert, d'analyse.

Les représentations sont assurées soit sous forme de listes (mots, paires, etc.), de matrices qui peuvent ensuite être transposées dans des logiciels d'analyse (TEWAT d'IBM, STATISTICA,... par exemple) ou dans des graphes du type Excel. La lecture est facilitée par des filtres accompagnés ou non de transcodages. Le reformatage peut être réalisé soit via des outils commerciaux (Infotrans) ou via des outils dédiés développés au CRRM. Les comparaisons dynamiques entre listes (provenant de serveurs ou des logiciels du CRRM) sont réalisables simultanément jusqu'à 32 listes à la fois).

Ces outils peuvent s'intégrer dans tout système de Veille Technologique ou d'Intelligence Stratégique, car ils permettent de répondre aux questions de base: qui fait quoi, où, quand, comment. Quelle est ma position comparée à..., dans quel réseau suis-je intégré.

Nous présentons en annexe la description des principaux logiciels réalisés dans ce cadre. Ces derniers acquis par des tiers, ou utilisés par nos élèves en cours de stages ou de thèses.

Bibliographie

Veille Technologique et Développement Industriel. De la Grande Entreprise aux PNOE et PMI.

Dou H.
Dunod, Paris,. 1995

Bibliometric Analysis of patent Documents for R&D Management

Quoniain L., Hassanaly, P., Baldit P., Rostaing H., Dou H.
4ième Congrès International, Bibliometry, Infometry, Scientometry, Berlin,
1993

Introduction à la scientométrie. De la bibliométrie à la Veille Technologique

Courtial J.-P.
Anthropos-Economica, Paris, 1990

Modelo de gestao de informacao estrategica de apoio a decisao

Dou H., Raimundo dos Santos
Caesura, 1993, p-84-87

La Science de l'Information

Le Coadic Y.
Que sais-je?, n° 2873, Paris

Apport de l'analyse factorielle et de la classification ascendante hiérarchique dans l'analyse des banques de données bibliographiques,

Paoli C., Billard P., Blanchet P., Longevialle C.,
Premier Colloque sur l'Information Ellaborée, Ile Rousse.- Edité dans la Revue Française de
Bibliométrie, numéro spécial, pp-65-75, 1987

Le brevet source d'information

Jakobiak F.
Dunod, Paris, 1994

Bibliometric analysis of patent documents for R&D management

Quoniam L., Hassanaly P., Baldit P., Rostaing H., Dou H.
Research Evaluation, pp. 13-18, Avril 1993

Le logiciel d'analyse de références et de dépêches d'agence SIADIC

Giraud E., Dou H.
Le Documentaliste, Mai 1994

Who wins from innovation

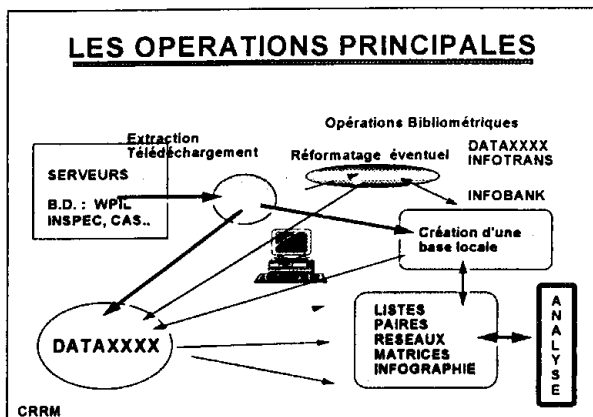
Jevons J
Technology Analysis and Strategic Management, 1993, p,399-412

Annexe

Traitement et Analyse de l'Information par la Méthode Intégrée du CRRM

La croissance constante des flux d'information ainsi que la multiplicité des sources a conduit le CRRM à mener depuis plusieurs années une réflexion au niveau des méthodes et des outils informatiques permettant le traitement automatique de l'information.

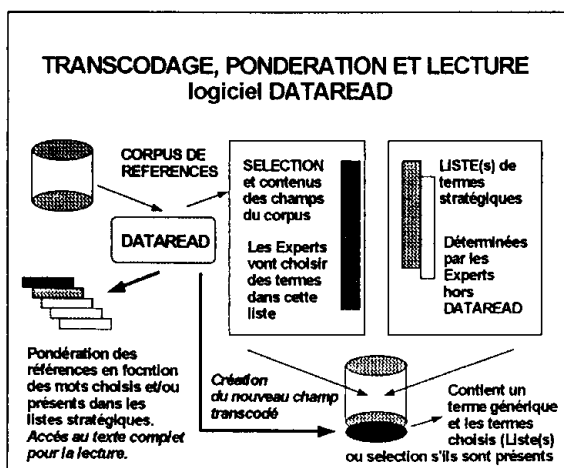
Les entreprises opposées à des concurrents organisés, répartis sur des zones géographiques de plus en plus larges et confrontées à des technologies évolutives doivent mettre en place des systèmes d'information élaborés intégrés à des services de Veille Technologique de plus en plus performants.



Pour faciliter l'action des entreprises dans ce domaine le CRRM a développé une méthode de traitement puissante s'appuyant sur des logiciels performants opérant sous MS/DOS et permettant une Veille Technologique répartie. Ces logiciels répondent au besoin de gestion de l'information, de l'analyse automatique de celle-ci, de la réalisation de corrélations simples ou très élaborées et de la réalisation de comparaisons puissantes ou de transcodages automatiques.

Répondre aux questions : qui fait quoi? où? comment? dans quel contexte constitue l'objectif que nous nous sommes fixé. La série des logiciels DATAxxxx, épaulés par quelques logiciels commerciaux et par les grands classiques de Windows (Word, Excel) apportent une réponse intégrée à ces questions.

Reformater, Gérer, Transcoder, Pondérer, Accéder aux informations



Ces étapes indispensables dans tout traitement automatisé sont réalisées par l'ensemble des logiciels suivants:

INFOTRANS et **INFOBANK** de la Société I+K avec laquelle nous collaborons depuis plusieurs années, apportent une réponse simple au reformatage et à la gestion des documents.

Des outils intermédiaires comme **DATACHANGE** permettent des opérations répétitives et simples de reformatage, lorsque les changements à effectuer ne sont pas très importants.

DATAREAD permet l'accès aux informations de façon pondérée à partir de listes stratégiques de termes ou à partir des

termes contenus dans un ou plusieurs champs (ou dans l'ensemble des documents) et présentés à

l'utilisateur qui sélectionnera les plus pertinents. Après pondération l'utilisateur peut accéder directement au texte complet du document.

DATAREAD apporte aussi une réponse très puissante aux opérations de trancodage d'un corpus de référence. Il permet soit à partir de listes stratégiques de termes, soit à partir des termes contenus dans un ou plusieurs champs (ou dans les documents complets), de créer dans chaque référence un champ supplémentaire contenant un terme générique choisi par l'utilisateur plus un ou plusieurs des termes présents dans la référence et sélectionnés par l'utilisateur ou issus de la liste des termes stratégiques.

Analyser un corpus, corréler les informations

Thèse de W. NIVOL - Application bibliométriques

LES POLYMERES CATIONIQUES
Matrice de correspondances
Types d'application - Propriétés - Fonctions chimiques

	A	B	C	D	E	F	G	H	I	TOTAL
TYPES D'APPLICATION										
COMPLEXE (644 desc)	86	28	38	18	6	4	18	18	11	284
EMULSIONS	56	28	14	27	18	18	18	18	8	228
COLORANTS	57	24	21	28	9	9	18	18	8	177
RESISTANCES	53	44	18	18	2	9	21	14	18	153
ANTICOLORANTS	84	21	18	12	3	12	12	8	8	208
DESCHROMES	28	18	6	27	8	15	18	8	23	153
PROPRIETES										
CONDENSATION	47	45	9	25	9	9	18	14	14	198
BOULERS	118	44	13	28	5	14	18	4	18	248
EFFET COPYANT	56	45	18	17	8	7	9	18	28	208
HYDRATATION	87	17	18	28	8	13	7	2	8	184
PROTECTION	67	27	31	22	8	12	18	18	8	211
POLYMERES EMULSIONS	37	63	7	18	7	7	18	11	18	211
BELLEANCE	33	28	8	18	18	14	14	5	24	211

A. REPERES DE D'ORIGINE E. COULEUR
 B. ADHESION CHIMIQUE F. PROPRETE
 C. COMPLEXE G. CHROMES
 D. CELLULOSE H. FAMILIARITE QUANTITATIVE

Valeurs de fréquences relatives à la somme de la colonne
 Valeurs de fréquences relatives à la somme de la ligne

CRRM

Quels sont les auteurs ou inventeurs principaux, les déposants les plus agressifs, quel est l'environnement d'une technologie, à quoi est-elle appliquée, où sont effectués les recherches et le développement, quelle est l'expertise d'un inventeur, d'un auteur, d'une institution? Quels sont les principaux réseaux de thèmes, d'auteurs, etc.?

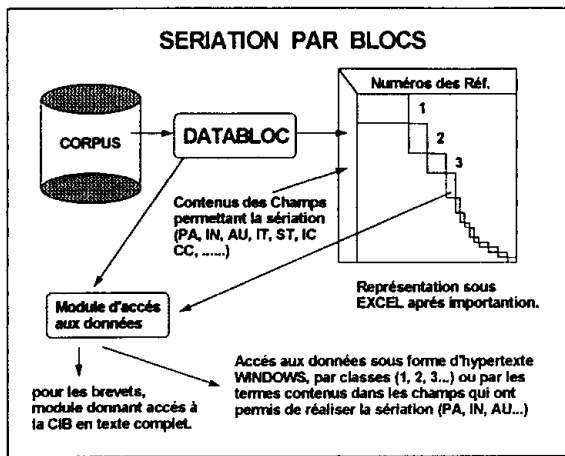
L'ensemble de ces questions trouvera une réponse par l'utilisation du logiciel **DATAVIEW**, puissant analyseur qui permet de constituer des listes, des matrices entre les éléments d'un ou de plusieurs champs, de sélectionner les gammes de fréquences à utiliser, les termes à

sélectionner. Les transpositions et importations dans des grapheurs (Excel) et/ou dans des logiciels d'analyse de données comme Clustan, Arcade, StatITCF, ... permettent des représentations infographiques puissantes ainsi que des calculs et simulations associés aux analyses.

Obtenir des présomptions, comparer de 2 à 31 listes simultanément

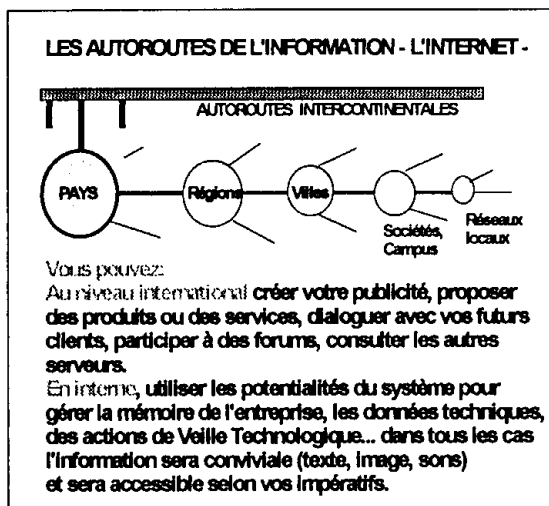
C'est une approche puissante de l'accès à l'information floue, sans avoir recours aux opérateurs booléens. La réponse à cette méthode de travail est fournie par le logiciel **DATALIST** qui permet par la comparaison de listes issues des commandes statistiques des serveurs ou issues des traitements de **DATAVIEW** d'apporter une facilité de travail incomparable dans tous les problèmes de «bench marking», que ce soit entre sociétés, institutions, chercheurs, etc... La comparaison simultanée des savoir-faire, des thèmes de recherche, par plusieurs dizaines de sociétés, d'individus ou d'institutions est un moyen particulièrement intéressant dans le domaine de la Veille.

La sériation par blocs



L'analyse statistique classique par la difficile représentation des corrélations dans l'espace doit souvent être complétée par une méthode de regroupement plus visuelle des données. Cette facilité est offerte par le logiciel **DATABLOC** qui permet par des méthodes de sériation avancées de regrouper les références à partir du contenu d'un ou de plusieurs champs documentaires. L'accès aux informations contenues dans les références regroupées par bloc (sériation par bloc) est réalisé automatiquement via l'hypertexte de windows, l'ensemble des données étant codées pour répondre à cet accès automatisé.

L'internet



Dans le cadre de nos activités nous avons développé une compétence dans le domaine de la gestion d'informations sous Internet et plus particulièrement dans le cadre du développement des serveurs WWW (World Wide Web) et de bases de données sous WAIS. Que ce soit ouvert sur réseau on en système interne, nous pouvons aborder avec vous la gestion de votre documentation interne, la gestion de la documentation technique, et de la mémoire de l'entreprise etc. Ces nouvelles méthodes de diffusion et d'accès à l'information peuvent être intégrées aux activités de votre société. Nous pouvons coopérer avec vous dans cette approche nouvelle et tester pour vous les potentialités de cette méthode pour accéder et diffuser les informations.

La réalisation de CD-ROM ou de présentations sous hypertexte

Ces nouvelles façons d'accéder à l'information sont particulièrement utiles dans le domaine de la Qualité, des Normes, des données techniques et de sécurité et de la gestion du patrimoine de l'entreprise. Nous possédons une certaine maîtrise dans ces domaines et nous avons développé une expertise dans divers secteurs. Vous pouvez nous consulter si vous avez à aborder de tels problèmes de gestion, stockage et diffusion de votre information. Que ce soit sous WWW ou Authorware ou pour gérer de simples données bibliographiques nous pouvons confectionner à la demande, un CD-ROM d'essai qui sera utilisable au niveau des tests ou de la validation avant l'impression définitive.

La formation

Enfin, il ne servirait à rien de posséder des outils performants s'ils n'étaient pas complétés par des méthodes d'utilisation et par un savoir-faire acquis après plusieurs années de recherche dans le domaine de la Veille Technologique. Ce transfert de savoir-faire sera réalisé par l'intermédiaire des formations dans l'entreprise ou au CRRM qui accompagnent tout achat de logiciels.

Nous vous proposons plusieurs façons d'acquérir les outils nécessaires à l'analyse de l'information: soit passer par des étapes intermédiaires consistant à faire réaliser des traitements à façon pour tester les produits, soit participer aux stages de formation de Veille Technologique que nous organisons, soit acquérir directement les logiciels et bénéficier des formations qui les accompagnent.

Ces formations peuvent être réalisées soit sous forme d'interventions ou de contrats de formation, soit dans le cadre légal de la formation continue.

Le CRRM

Le CRRM est un laboratoire d'Enseignement et de Recherche qui s'est développé et spécialisé dans les traitements et l'analyse de l'Information, dès 1978. Le développement d'enseignements spécifiques dans ce domaine nous a conduit à la mise en place progressive de la **Filière Information pour l'Entreprise**.

Les activités internationales du CRRM sont nombreuses et orientées spécifiquement vers un certain nombre de pays: Luxembourg, USA, Canada, Japon, Brésil, Indonésie.

Une forte composante industrielle est toujours présente dans notre activité. En effet, soit sous forme de stages, soit sous forme de conventions CIFRE, soit sous forme de contrats nous coopérons avec de nombreuses industries petites, moyennes ou grandes.

Au plan de la recherche le CRRM développe des méthodes et des concepts nouveaux en Veille Technologique et dans le domaine de l'Information Stratégique. *Cet ensemble s'appuie aussi sur des logiciels spécifiques bénéficiant de l'ensemble des avancées méthodologiques et conceptuelles de plus de 15 années de réflexion et de recherche.*

Enfin, il est possible au CRRM de travailler sous forme de Conseil ou par l'Intermédiaire de contrats de collaborations avec les collectivités locales et nationales et avec l'industrie. Très souvent ces activités sont couplées avec des travaux d'étudiants réalisés sous forme de stage, de mémoire, ou de conventions CIFRE.

Le CRRM a mis au point un ensemble de conventions au secret qui lient ses membres ou les étudiants travaillant sur un sujet précis avec l'industriel. Les soutenances mémoires, thèses, peuvent être réalisées dans des conditions de sécurité et de confidentialité draconiennes si cela est nécessaire.

Pour plus d'information, vous pouvez contacter le CRRM au 91288050 ou 91288697 ou faxer au 91288712 ou 91288749 ou écrire:

CRRM, Université Aix-Marseille III,
Centre Scientifique de St Jérôme 13397 Marseille Cedex 20
crrm@crrm.univ-mrs.fr