

## Index des publications VSST'98



[La norme AFNOR XP X 50-053 et la pratique de la veille](#) H.Stiller

[Théorie et pratique de la veille: quelques retours d'expérience contribuant à l'émergence du concept d'intelligence stratégique](#) H.Lesca, S.Bianco

[Méthodes de structuration pour l'analyse stratégique des univers scientifiques: les techniques de citation](#) M.Zitt, E.Bassecoulard

[Définition d'un dispositif de veille stratégique pour les PME tunisiennes](#) S.Chouk-Kamoun, M.Salles

[PME et marché international: une approche exploratoire des problèmes de traitement de l'information](#) M.Boutary

[Les PME et l'intelligence économique: la synergie public-privé](#) P.Jacques-Gustave, N.Moinet

[Méthode de conception de produits d'intelligence économique destinés à des PME](#) M.Salles, T.Zid

[SIMBAD: système d'interrogation multibases d'aide à la décision](#) P.Baldir, V.Fabreguettes, A-M.Jonquière

[Optimisation du choix de la terminologie pour la reformulation de requêtes: cas des multi-termes](#) B.Dousset, S.Kanoun

[L'analyse de données textuelles: nécessité de l'aide à la préparation des données. L'expérience d'Aérospatiale](#) S.Artillan, M.Lalaude, M.Meyer

[Vers un nouveau mode d'interrogation des documents issus du web](#) F.Riahi

[Indexation sur Internet, metadata, Dublin Core](#) J.Ducloy

[Analyse d'informations issues du web avec Tétralogie](#) T.Dkaki, B.Dousset, J Mothe

[Brevet et innovation: une méthode de recherche de nouvelles applications pour un matériau](#) F.Guesdon, P.Hassanaly

[Veille technologique: adaptation à un système d'information économique tunisien](#) F.Chichti, P.Hassanaly

[Stratégie technologique: applications à l'industrie du raffinage du pétrole au Mexique](#) C-E.Escobar-Toledo, R.Cassaigne-Hernandez

[Extraction automatique et représentation graphique de données biologiques: les interactions génétiques moléculaires](#) V.Pillet, B.Roudani, L.Quoniam, B.Jacq

[Contribution à la définition d'un vigiciel: quelle modélisation de l'information factuelle, événementielle et référentielle ?](#) D.Graveleau, L.Berti

Analyse de données et analyse de mots associés, comparaison d'algorithmes différents sur un corpus concernant la prise en compte du risque dans le développement des OGM dans le domaine des végétaux M-A De Looze, A Roy, M Reinert, O Jouve, R Coronini

Un outil pour la veille: le maillage technologique de références bibliographiques scientifiques P.Faucompre, L.Quoniam

Enseignement tiré d'un programme de compréhension automatique de comptes rendus médico-techniques M.Roux, V.Ledoray

Système 3AD: un outil de classification à caractère linguistico-mathématique I.Timimi

Les processus d'apprentissage au coeur de la veille technologique dans un environnement de R&D C.Rondeau

Organisation et gestion des connaissances en veille scientifique et technologique X.Polanco, C.Francois, J.Royaute, L.Grivel, D.Besagni, M.Dejean, C.Oto

Création d'hypertextes automatiques appliqués à la veille V.Leveille, H.Rostaing, L.Quoniam

Une méthode de détection des signaux faibles: application à l'émergence des dendrimères C.Roux, B.Dousset

# CREATION D'HYPERTEXTES AUTOMATIQUES APPLIQUES A LA VEILLE

LEVEILLE V., ROSTAING H., QUONIAM L.

## CRRM

Centre scientifique de Saint-Jérôme

Av. Escadrille Normandie-Niemen

13397 Marseille Cedex 20

Tel : 04 91 28 87 40 Fax : 04 91 28 87 12

e-mail : {leveille, rostaing, quoniam}@crrm.univ-mrs.fr

Un monde de plus en plus complexe, et ouvert aux innovations, peut offrir de multiples opportunités à ceux qui savent anticiper mais peut également noyer des entreprises qui n'auront pas su saisir à temps ces opportunités.

Dans l'environnement actuel, seule une attitude appropriée de veille est susceptible de déjouer les pièges qui menacent les entreprises de paralysie, car la mission de la veille est précisément de discerner dans une multitude d'informations les signaux faibles, mais pertinents qui préfigurent l'avenir.

Cette information est généralement facile d'accès mais surabondante, fragmentaire et donc peu sûre. Elle se présente sous multiples formes : textuelle (références bibliographiques, articles de presse, rapports d'experts), numérique et infographique (résultats d'analyses bibliométriques, tableaux de bord et tableaux prévisionnels), d'image (plaquettes publicitaires ...). Trier, recouper et synthétiser l'information devient alors essentiel pour la rendre fiable et efficiente. Pour cela, il est nécessaire de pouvoir favoriser ces recoupements quel que soit son format sous une même interface avec un outil destiné aux utilisateurs du système d'information.

L'objet de cette communication est de présenter l'élaboration au CRRM, d'un outil de navigation relationnelle. Cet outil, basé sur le principe de la construction automatique hypertexte, propose aux utilisateurs du système de veille les moyens pour analyser, trier et recouper l'information.

## Introduction

Le processus de veille technologique peut se résumer à ces trois phases : collecte de l'information, analyse de l'ensemble des informations collectées, diffusion de cette analyse pour action.

L'une des principales difficultés de la seconde étape concerne le volume important d'informations à analyser, recouper, synthétiser en un temps très court. Veilleurs et experts du domaine interviennent ensemble sur cette phase essentielle en veille technologique. Le premier est chargé de préparer le dossier de veille à l'analyse de l'expert. En organisant et en structurant le dossier, le veilleur s'assurera ainsi de la pleine collaboration de l'expert.

L'organisation du dossier de veille doit permettre à l'expert [ROUSSEAU 97]:

- | D'avoir une vision globale du dossier
- | De positionner son entreprise, son activité ou ses concurrents dans le dossier
- | D'explorer la masse de documents mise à sa disposition. L'exploration est une phase indispensable en veille. L'expert doit pouvoir "entrer" dans le dossier par une information, un mot-clé ou un auteur qu'il connaît et pouvoir connecter d'autres informations qui lui sont inconnues.

Les outils bibliométriques classiques réalisent d'ores et déjà ces opérations de cartographie et de structuration du dossier [WHITE 89], [DOUSSET 95].

L'exploration et la navigation au sein de la masse de documents relève d'avantage du domaine des outils de gestion de l'information que des outils bibliométriques. On peut distinguer deux approches :

- | L'approche booléenne. L'utilisateur interroge le fonds documentaire par des requêtes booléennes ou en langage naturel. Cette approche requiert de la part de l'utilisateur une certaine connaissance de la base.
- | L'approche hypertextuelle. Elle permet à partir d'un document de départ de naviguer au sein de l'ensemble des documents sans avoir à priori une parfaite connaissance de sa structure.

Cette approche hypertextuelle est très séduisante en veille technologique par sa capacité à produire de l'information endogène : information non présente dans le dossier mais issue du croisement de deux informations.

Deux freins majeurs ont retardé la généralisation de l'utilisation d'hypertextes en veille [LELU 95]:

- Les limites liées à l'édition manuelle des liens. D'une part, lier manuellement deux documents impose une bonne connaissance du contenu de la base. D'autre part, le temps utilisé pour la création manuelle serait inacceptable dans un processus de veille.
- Le manque de vision globale. Le manque de vue d'ensemble du corpus est à l'encontre de ce qui est requis d'un outil de veille. L'hypertexte classique présente uniquement des liens entre documents, sans outils de cartographie du corpus.

En utilisant des à des nœuds d'hypertexte à rôle de synthèse, la structuration des informations présentée par A. Lelu, permet d'obtenir une vision globale des documents. Cette méthode de construction, développée tout d'abord par Damashek [Damashek 95], est basée sur une analyse classificatoire, les nœuds de synthèse représentant un ensemble de documents.

D'autres méthodes ont également été développées, notamment la méthode Tachir [AGOSTI 94]. Cette méthode n'est pas basée sur l'analyse classificatoire mais établit les liens entre documents via des termes indexés (mot-clé, auteurs...).

En 1995 [ROSTAING 1995], le CRRM a évalué l'utilisation de la navigation hypertextuelle dans le cadre de l'analyse d'un dossier de veille. Une analyse statistique relationnelle des données, a permis de créer des nœuds de synthèse en regroupant par thèmes les différents documents du dossier.

## **Construction d'hypertextes automatiques appliqués à la veille**

Nous développons au CRRM un outil informatique offrant une navigation issue conjointement de l'hypertexte et de la bibliométrie. Cet outil, de part cette filiation commune, permet de proposer à l'utilisateur une vision à la fois synthétique et globale du dossier, tout en conservant l'accès à l'information primaire.

Pour bâtir notre hypertexte, nous avons choisi de tirer profit de la structure relationnelle des données manipulées en veille. En effet, les champs auteurs extraits de documents issus des bases de données, permettent de relier deux documents issus de la même base mais aussi de deux bases différentes. Associer deux documents provenant de deux bases de structures différentes est très important puisqu'un dossier de veille technologique doit être constitué à partir d'informations de sources différentes. Il est, par exemple, très important de pouvoir associer les articles scientifiques écrits par un auteur avec les brevets déposés par celui-ci.

De plus, il nous est apparu important de laisser l'utilisateur créer son propre parcours sans l'influencer par des liens préétablis. Les liens sont créés automatiquement. L'utilisateur en fonction de son domaine, de ses pôles d'intérêt, de ses aspirations choisira de développer une voie plutôt qu'une autre, en n'étant jamais influencé au cours de sa navigation par des liens prédéfinis.

Cependant, nous nous sommes très rapidement heurtés au problème de la multiplication des liens entre documents, ce qui perturbait la vision globale offerte à l'utilisateur. En effet, les termes triviaux à haute fréquence, liaient tous les documents et les termes à faible fréquence parasitaient, en quelque sorte, le schéma de navigation en ne liant que très peu de documents.

### **Utilisation de filtres de fréquences appliqués à la construction automatique d'hypertextes**

Pour résoudre ce problème, nous avons choisi de proposer à l'utilisateur un filtre de fréquence, suggérant d'éliminer de la navigation, d'une part les termes triviaux et d'autre part les termes marginaux. Un calcul sur la méthode d'entropie [LHEN 95] propose à l'utilisateur une répartition automatique des termes en trois zones de fréquence. Chaque terme (auteurs, mot-clé, code de classification documentaire..), est qualifié comme appartenant à la catégorie HF (haute fréquence, termes triviaux), MF (moyenne fréquence) ou BF (basse fréquence, termes marginaux). Les termes qualifiés MF peuvent, par exemple, être sélectionnés par l'utilisateur pour associer deux documents. Cette méthode nous permet ainsi "d'alléger" la représentation hypertextuelle et d'offrir ainsi une meilleure vision du dossier à l'expert. Bien évidemment, le logiciel propose à l'utilisateur d'ajuster à la suite de son expertise les seuils de ces trois zones.

## **Navigation au sein du dossier**

En outre, le mode de présentation de l'outil offre la possibilité de visualiser les liens entre documents, mais aussi entre termes et documents ainsi qu'entre termes. Par exemple, le logiciel montre graphiquement quels sont les mots-clés associés à tel inventeur ou à telle société. En utilisant un hypertexte classique, pour connaître quels termes sont associés à l'auteur, il est nécessaire pour cela de parcourir le document. Notre logiciel permet très rapidement d'obtenir ce renseignement tout en conservant l'accès à l'information primaire. L'expert a ainsi une vision globale du dossier, tout en ayant les moyens de recouper, de croiser les informations en naviguant au sein du dossier.

## **Conclusion**

En répondant aux besoins d'analyse des utilisateurs du système de veille, nous proposons un outil informatique adapté à la veille offrant à la fois une vision synthétique et globale à l'utilisateur ainsi que des possibilités de navigations objectives au sein du dossier de veille. Ces différentes fonctionnalités facilitent ainsi l'expertise du dossier en favorisant le passage de l'information brute à l'information élaborée.

## Bibliographie

[AGOSTI 1994] : AGOSTI M., MELUCCI M., CRESTANI F.; "TACHIR : a tool for automatic construction of hypertexts for information retrieval", *RIAO 94 Intelligent multimedia information retrieval systems and management*, Rockefeller University New-York, N.Y.-USA, October 11-13, 1994, pp. 338-357

[DAMASHEK 1995] : M. DAMASHEK; "Gauging similarity with n-grams : language-independent categorization of text", *Science*, vol. 267, février 1995

[DOUSSET 1995] : DOUSSET B., "Le logiciel d'études bibliométriques TETRALOGIE de l'IRIT", *Colloque VSST'95 Veille stratégique, scientifique et technologique*, Toulouse, 25-27 octobre, 1995, pp. 431-471

[LELU 1995] : LELU A; "Hypertextes : la voie de l'analyse de données", *Troisièmes Journées Internationales de l'Analyse des Données Textuelles*, Rome, 13-15 décembre 1995, pp. 85-96

[LHEN 1995] : LHEN J., LAFOUGE T., ELSKENS Y., QUONIAM L., DOU H., "La "statistique" des lois de Zipf", *Colloque Les journées d'information élaborée*, Ile Rousse, 1995, pp. 135-146

[ROSTAING 1995] : ROSTAING H., DJAOUZI S., LA TELA A., AVIGNON T., QUONIAM L., "Analyse bibliométrique multi-bases pour l'élaboration d'un dossier électronique de veille technologique", *Colloque VSST'95 Veille stratégique, scientifique et technologique*, Toulouse, 25-27 octobre, 1995, pp. 153-168

[ROUSSEAU 97] : ROUSSEAU F., THIL J., "Veille et informatique : des besoins aux solutions", *Technologies Internationales*, n° 39, novembre 1997, pp. 33-36

[WHITE 1989] : WHITE H. D., MC CAIN K. W., "Bibliometrics", *Annual Review of Information Science and Technology (ARIST)*, vol. 24, pp. 119-186, 1989