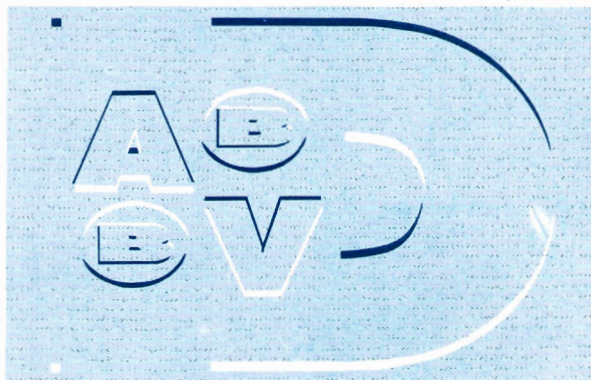


Cahiers de la documentation Bladen voor de documentatie

JUSTIFICATIE



ASSOCIATION
BELGE DE
DOCUMENTATION

BELGISCHE
VERENIGING VOOR
DOCUMENTATIE

N° 1 - 1992

ISSN 0007-9804

Editeur responsable,

Verantwoordelijke uitgever,

PAUL HUBOT

c/o Centre de Recherches Routières - Boulevard de la Woluwe, 42 - 1200 Bruxelles

DÉPOSÉ A: 1160 BRUXELLES 16
TRIMESTRIEL - DRIEMAANDELIJKS

CODES DE CLASSEMENT ET INFORMATION STRATEGIQUE

OUTILS DE TRAITEMENTS AUTOMATIQUES

Stéphane DUMAS, Henri DOU, Luc QUONIAM
 Centre de Recherche Rétrospective de Marseille
 Faculté des Sciences et Techniques de St Jérôme
 Avenue Escadrille Normandie Niemen, 13397 MARSEILLE CEDEX 13

L'accroissement rapide de la compétitivité, le développement des technologies et les menaces qu'ils font peser sur la survie des entreprises ont conduit leurs dirigeants à considérer le management des ressources technologiques comme un axe majeur de leur stratégie. L'importance accordée par les décideurs aux informations technologiques a entraîné une mutation de la fonction documentaire. Il devient nécessaire de fournir une information à plus haute valeur ajoutée sous des formes différentes de celles du passé. Ces améliorations ont été grandement favorisées par le développement des stations de travail qui permettent aujourd'hui de formaliser et d'analyser les résultats sous des aspects plus propices à la prise de décision.

BIBLIOMETRIE - INFOGRAPHIE : DES OUTILS D'AIDE A LA DECISION

La bibliométrie, qui devient une composante importante des systèmes d'aide à la décision tels que la veille technologique, s'est trouvée au faite de ces modifications. Elle recouvre un ensemble de méthodes statistiques d'évaluation de l'information scientifique et technique que A. PRITCHARD (1) définissait dès 1969 comme :

"L'application des mathématiques et des méthodes statistiques aux livres, articles et autres moyens de communication."

Les masses d'informations à traiter sont souvent trop importantes pour pouvoir réaliser une analyse manuelle ou par simple lecture séquentielle. L'ensemble des méthodes bibliométriques permet de condenser ces informations de manière à rendre les résultats exploitables. Ces résultats se présentent en premier lieu de manière tabulée mais si cette forme est la seule à garantir l'intégralité du contenu informationnel des résultats, son exploitation est souvent fastidieuse et sa compréhension au niveau global n'est pas toujours aisée. Afin de faciliter l'étude des résultats, on a pris l'habitude de les présenter sous forme infographique, les rendant ainsi plus directement lisibles (2).

L'infographie englobe un ensemble de techniques qui concernent la représentation iconographique automatique d'un lot d'informations. Ces méthodes proposent une globalisation des résultats utiles aux décideurs. Ainsi, dans la plupart des études bibliographiques, l'infographie se présente comme une grille de lecture de documents primaires, un support visuel accompagnant une argumentation. Cependant, si leur caractéristique principale est de permettre d'étayer un argument, leur interprétation reste souvent du domaine des spécialistes.

LES CODES : SOURCE D'INFORMATIONS STRATEGIQUES

Une classification est un instrument de travail qui permet de représenter de manière synthétique le sujet d'un document et de regrouper des ouvrages (ou des articles) par affinité de contenu. Le codage est la matérialisation numérique, alphabétique ou alphanumérique de la classification (3). La naissance d'une codification simple de l'information sous forme de partitions globales a permis de réaliser une fragmentation de l'information dans de multiples domaines (chimie, physique, économie, brevets, ...). Tout d'abord réalisée à des fins documentaires, elle est devenue une aide significative au niveau du traitement bibliométrique. En effet, les codes autorisent le regroupement des ouvrages (ou articles) voisins et leur recherche par thèmes. Ils sont abstraits mais obéissent nécessairement à une logique puisqu'ils reflètent la parenté du contenu des documents. Ils sont sans rapport avec les termes des langues naturelles et offrent la possibilité de s'affranchir des notions de langage spécifique ainsi que des évolutions du langage dans le temps car ils restent fixes pour des périodes relativement importantes. En outre, les codes sont attractifs, ils vont permettre de réaliser facilement des représentations infographiques ou en réseau car leur nombre est fini et beaucoup plus faible que tous les concepts qu'ils englobent. Indicateurs de tendances, critères d'évaluations, ils permettent de se faire une opinion sur un domaine particulièrement flou où les hiérarchies sont souvent liées à l'habitude.

L'étude des champs codes issus de l'interrogation d'une base de données doit être soumise à quelques précautions inhérentes à divers facteurs qui sont :

- * la finalité des codes : les classifications sont conçues pour regrouper les notices par affinité de contenu, elles ne sont en aucun cas créées à des fins analytiques. Ce sont les personnes désirant les utiliser dans ce but qui trouvent dans les codes un support d'étude. Il paraît donc évident que l'analyse dépend de la conception de la classification et que divers codes ne s'étudient pas de la même manière, certains demeurant par ailleurs non analysables car n'offrant pas des degrés de hiérarchie suffisamment précis,
- * le besoin d'exhaustivité : ces méthodes donnent des vues d'ensemble et mettent en évidence des fluctuations, des orientations, des affinités. Elles doivent par conséquent prendre en compte la totalité des données. Il faut donc avoir recours à des banques de données les plus exhaustives possibles et effectuer des analyses répétitives afin de déterminer les évolutions,
- * l'indexation: il s'agit là d'un problème majeur, l'indexation variant considérablement selon les bases de données et les origines de leurs analyseurs. On comprend donc la nécessité de focaliser une étude sur une seule base si l'on désire conserver la légitimité d'un traitement car il est dangereux de raisonner de la même manière face à des conceptions sémantiques et des niveaux de compétences différents,
- * l'évolution des plans de classement au cours du temps : une étude pour être valable doit porter sur des périodes recouvertes par des plans de classement identiques.

Si l'analyse des codes nécessite une approche circonspecte, il convient d'insister sur les nombreux avantages que recouvre une démarche de ce type et qui sont notamment de (4, 5, 6) :

- * permettre une saisie minimale,
- * hiérarchiser, si nécessaire, l'information donnant ainsi une signification de plus en plus poussée au code suivant sa longueur (codes C.I.B., Inspec, ...),
- * présenter les résultats sous des formes agréables, dynamiques,
- * permettre une visualisation globale, explicite,
- * s'affranchir des problèmes de modes linguistiques,
- * mettre en évidence la situation en volume des travaux et leur répartition,
- * décrire et comparer les travaux d'année en année,
- * mettre en évidence des profils,
- * évaluer des tendances, des évolutions,
- * donner une image des liaisons passées, des orientations à prendre,
- * mesurer l'effet des industries, des politiques de recherche,
- * décrire des réseaux de relations scientifiques entre différents domaines,
- * représenter et comparer des taux de croissance,
- * replacer un sujet d'étude dans un contexte élargi,
- * dégager des axes de développement.

Nous pouvons donc observer que l'analyse et l'interprétation des codes pourra permettre de dégager des synergies à créer pour développer une (des) activité(s) précise(s) et d'établir les planifications adéquates. Ils se présentent en cela comme des outils de veille performants. C'est la raison pour laquelle, le Centre de Recherche Rétros-

pective de Marseille, développe depuis plusieurs années, des méthodes de traitement des codes afin d'aboutir à des indicateurs réels, permettant d'amorcer les discussions indispensables liées à la programmation de la recherche.

UN OUTIL DE TRAITEMENT DES CODES : DATACODE

Parmi les outils élaborés dans ce but, nous retiendrons le logiciel DATACODE capable de traiter les codes chemical abstracts issus des bases CAS82/CAS77/CAS72/CAS67, les codes inspec issus des bases INSPEC/INSP69/76, les codes de la classification internationale des brevets, les derwent classes et les manual codes issus des bases WPI/WPIL. Ce logiciel dont les premiers travaux ont été réalisés par le professeur H. DOU est aujourd'hui développé en BASIC 7.1 Professional Development System @MICROSOFT. Ce BASIC intègre toutes les commandes du Quick Basic 4.5 et apporte des améliorations considérables telles que la gestion de la mémoire paginée (EMS), des vitesses d'exécution nettement plus élevées, la gestion des fontes graphiques.

La méthodologie est basée sur l'interrogation d'une des bases précédentes via le serveur ORBIT INFORMATION TECHNOLOGIES, le téléchargement d'un corpus - technique qui connaît un essor constant et qui désormais est pleinement reconnue (7, 8) - et son traitement en mode local. Les résultats se présentent sous différentes formes :

- * un fichier résultat présentant les libellés des codes concernés, les fréquences d'apparition, les fréquences des paires,
- * un graphe des paires,

- * une représentation cartographique en trois dimensions des fréquences d'apparition.

Les codes traités ont été choisis en fonction de la qualité d'indexation et de la couverture exhaustive des bases qui les produisent, cependant, les représentations obtenues et les conclusions qui pourront en découler seront fonction du caractère intrinsèque de ces codes. En effet, la structure des plans de classement diffère d'une base à une autre entraînant ainsi une modification des possibilités de traitement. Ainsi, si nous considérons les codes Inspec nous pouvons remarquer qu'ils sont hiérarchisés (la définition du domaine est d'autant plus précise que le nombre de digits est grand) et pondérés (les sections qui correspondent aux domaines principaux du travail analysé - sections principales - sont affectées d'une étoile) :

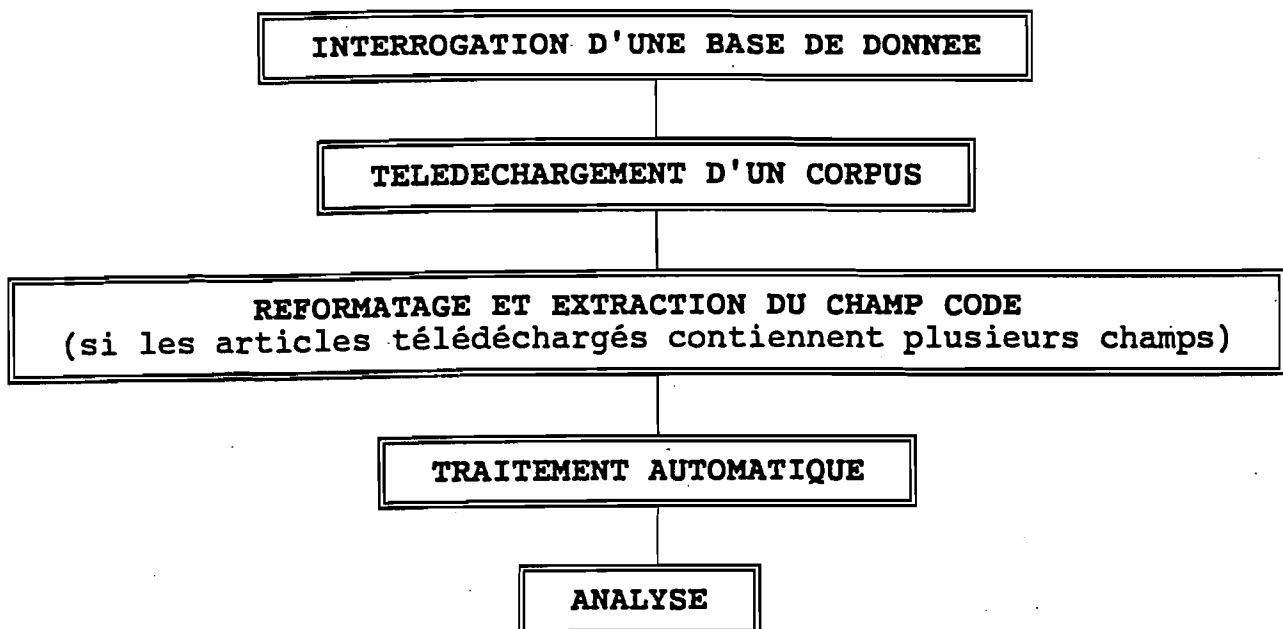
*champ code Inspec : CC - *A9385;
A9135D; A0720; *B7710; B7230C*

à l'inverse, les codes C.I.B. ne font pas apparaître de pondération mais seulement une hiérarchisation liée au nombre de digits :

*champ code C.I.B. : IC - G30G-005/04;
G03G-013/22*

Chaque champ code présente donc ses propres particularités multipliant ainsi les manières d'aborder une étude suivant la précision recherchée ou les corrélations que l'on désire mettre en évidence. La conception de ce logiciel a donc nécessité un examen approfondi de la structure des plans de classification des bases considérées et des différentes combinaisons analytiques réalisables, au même titre, son utilisation suppose une connaissance préalable du type de code que l'on étudie afin de ne pas tirer de conclusions trop hâtives et de ne pas s'investir dans des recherches sans fondement.

Nous pouvons résumer la démarche méthodologique par le schéma suivant :



EXEMPLE DE TRAITEMENT

Les résultats proposés sont relatifs à l'étude d'un échantillon de la base CHEMICAL ABSTRACTS SERVICE 82 dans le domaine de la chimie analytique. Le nombre de références traitées est de 594. Nous tenons à souligner que nous ne prétendons pas réaliser ici une étude

de précise en cherchant à dégager des conclusions stratégiques, les représentations présentées doivent donc être considérées comme de simples illustrations visant à démontrer le potentiel de l'outil utilisé.

Exemple de notice Chemical abstracts

AN - CA96-188185(22)
 TI - Calorimetric investigation of the interactions between oxygen and hydroxyl groups in (alcohol + ether) at 298.15 K
 AU - Villamanan, M. A.; Casanova, C.; Roux, A. H.; Grolier, J. P. E.
 SO - J. Chem. Thermodyn. (JCTDAF), V 14 (3), p. 251-8, 1982, ISSN 00219614
 OS - Univ. Clermont II, Lab. Thermodyn. Cinet. Chim., Aubiere, Fr., F-63170
 DT - J (Journal)
 LA - Eng
 CC - SEC69; SEC22; SEC23
 IT - 60-29-7, prop; 111-96-6; 142-96-1; 628-28-4; 629-14-1; Ethers, prop, aliph.: (heats of mixing of, with alcs.)
 IT - 64-17-5, prop; 67-56-1, prop; 71-23-8, prop; 111-27-3, prop; Alcohols, prop, aliph.: (heats of mixing of, with ethers)
 IT - Chains, chemical: (lengths of, of alcs. end ethers, heats of mixing in relation to)
 IT - Heat of mixing: (of alcs. and ethers, chain length in relation to)
 ST - alc; ether; heat; mixing; methanol; ether; heat; mixing; ethanol; ether; heat; mixing; propanol; ether; heat; mixing; heptanol; ether; heat; mixing; butyl; ether; alc; heat; mixing; ethyl; ether; alc; heat; mixing; methyl; butyl; ether; alc; mixing; heat; dioxaoctane; alc; mixing; heat; trioxaoctane; alc; mixing; heat

Notons pour une meilleure compréhension, que les codes C.A.S. sont au nombre de 80 et se répartissent en 5 grands groupes (Section Grouping) :

* biochemistry sections (codes 1 à 20)

* organic chemistry sections (codes 21 à 34)

* macromolecular chemistry sections (codes 35 à 46)

* applied chemistry and chemical engineering sections (codes 47 à 64)

* physical and analytical chemistry sections (codes 65 à 80)

Analyse des paires de codes

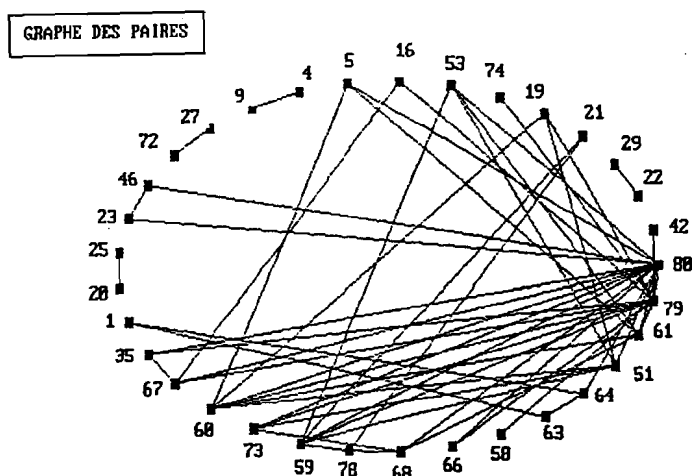


FIG 1

L'analyse conduit à définir des relations entre des pôles de recherche sans tenir compte de leur importance relative. Nous mettons donc en évidence des relations entre pôles sans se préoccuper de la hiérarchie (les codes C.A.S. faisant apparaître des sections principales et secondaires relatives à l'importance des domaines abordés dans un

item). Ainsi dans le cas présent, nous observons que certains pôles forment des cliques ouvertes : 20-25, 27-72, 4-9, 22-29, alors que d'autres sont très fortement liés et forment de véritables réseaux de relations : 80, 79, 61, ... Enfin, nous remarquerons le pôle 42 qui quoique lié au pôle 80 ne développe pas d'autres relations.

Analyse des 5 section grouping

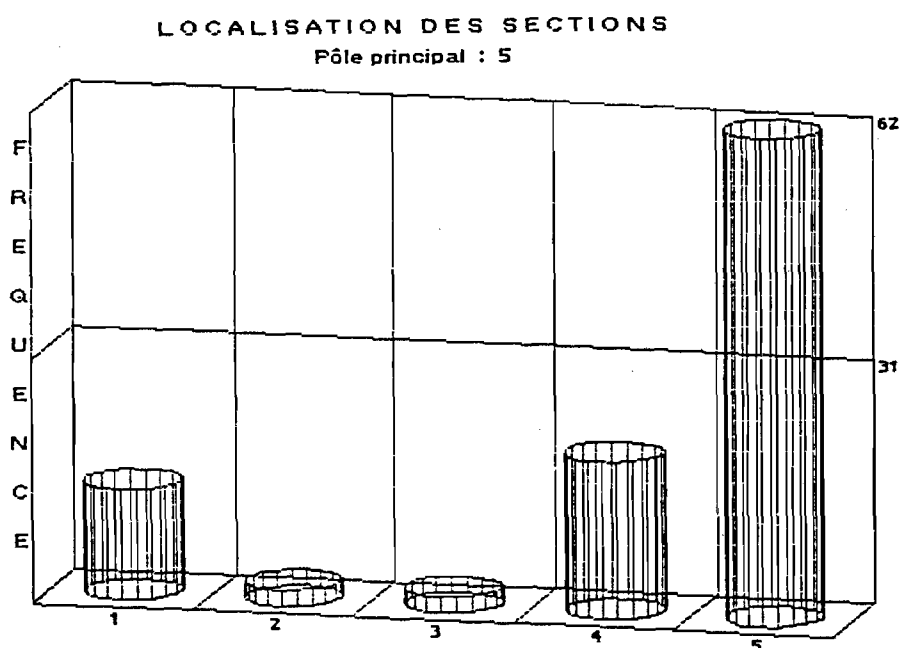


FIG 2

Nous pouvons remarquer la très forte proportion du groupe physical and analytical chemistry démontrant que les domaines de la physique et de la chimie analytique

ont fait l'objet de nombreuses études alors que les domaines de la chimie organique et macromoléculaire ont été peu investis.

Analyse des 80 sections

LOCALISATION DES SECTIONS
Pôle principal : 79

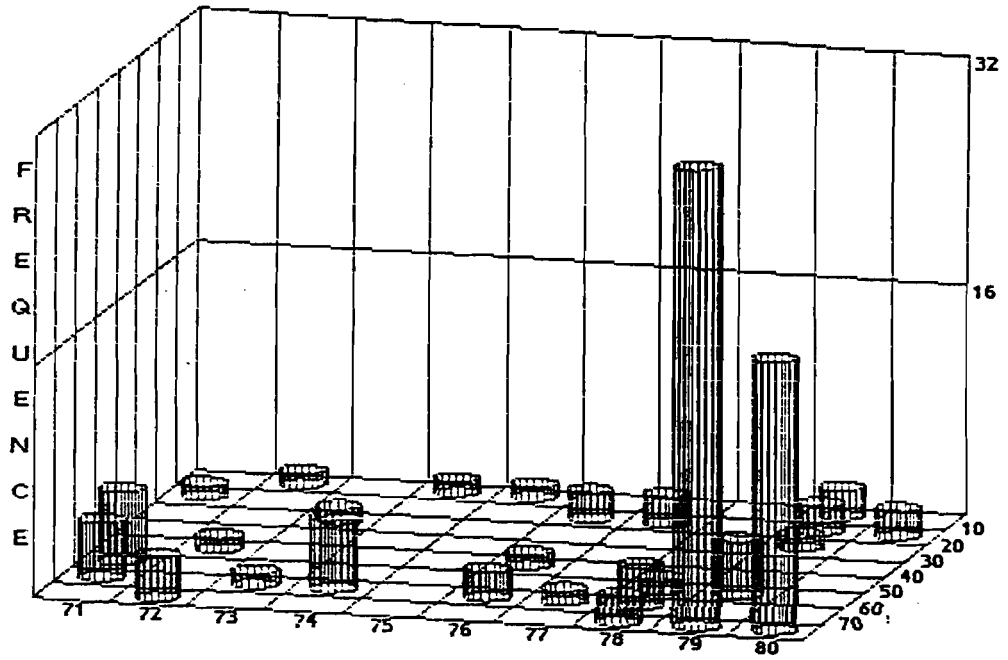


FIG 3

LOCALISATION DES SECTIONS
Pôle principal : 79

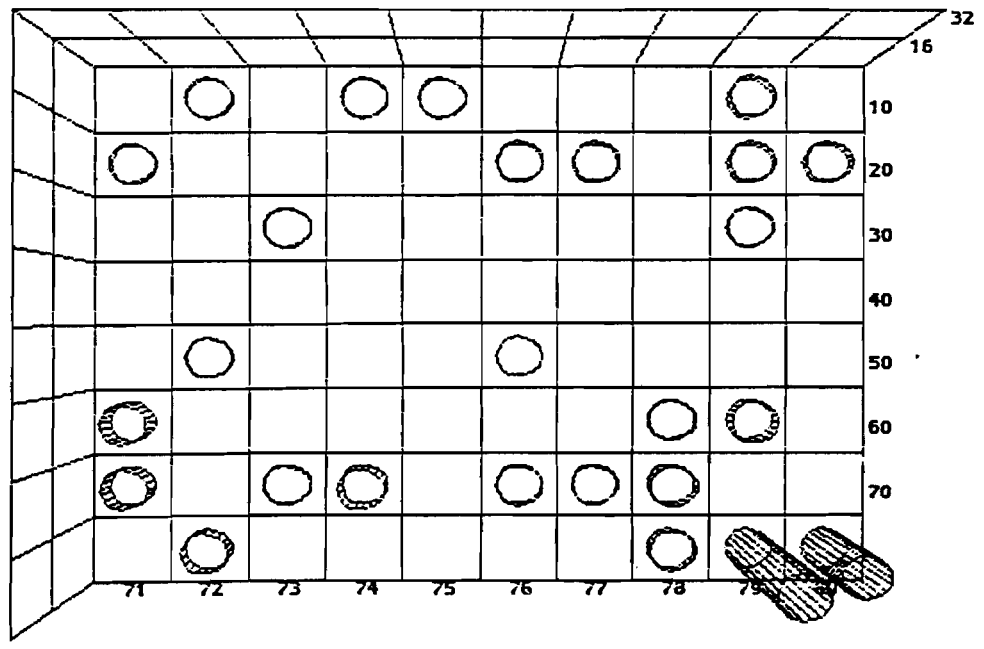


FIG 4

L'apport de la troisième dimension dans la conception graphique, offre une nette amélioration des possibilités d'observation car s'il est nécessaire de figer une représentation pour pouvoir la diffuser, l'aspect dynamique permettra de trouver l'angle le plus évocateur. De plus, la 3-D confère à cet outil un dynamisme susceptible d'éveiller la curiosité des utilisateurs. C'est un aspect sur lequel il convient d'insister, car, dans une approche de sensibilisation, le caractère ludique des outils utilisés s'avère être un facteur déterminant.

Note : les auteurs remercient Orbit Search Service pour l'aide apportée lors de la réalisation de ce travail.

BIBLIOGRAPHIE

- (1) PRITCHARD A., "Bibliometrics" - Volume 1, 1981.
- (2) DUTHEUIL C., "Infographie et contenu informationnel en bibliométrie"/Congrès de l'île Rousse, S.F.B.A., 1991.
- (3) MENIEZ J., "Les langages documentaires et classificatoires"/Les éditions d'organisation.
- (4) DOU H., HASSANALY P., QUONIAM L., "Les publications en chimie - les indicateurs de tendances"/Revue française de bibliométrie, déc. 1989.
- (5) DOU H., HASSANALY P., QUONIAM L., LA TELA A., "Veille Technologique et information documentaire - De l'usage de la bibliométrie dans les services de documentation"/Documentaliste, vol 27 n° 3, mai-juin 1990.
- (6) DOU H., HASSANALY P., LA TELA A., QUONIAM L., "Advanced interfaces to analyse automatically online database set of answer"/Elsevier science publisher, 1991.
- (7) LARDY J.P., "Connecter un micro-ordinateur à un serveur : comment, pourquoi ?"/Documentaliste, vol 23 n° 2, mars-avril 1986.
- (8) LARDY J.P., BADOR P., "Les techniques de téléchargement et le téléchargement de quelques logiciels de communication"/Documentaliste, vol 27 n° 2, mars-avril 1990.