



HAL
open science

La veille au Cemagref: une réalité quotidienne et un objectif commun

Raymond Schiano

► **To cite this version:**

Raymond Schiano. La veille au Cemagref: une réalité quotidienne et un objectif commun. Sciences de l'information et de la communication. 1997. hal-03177594

HAL Id: hal-03177594

<https://hal.inrae.fr/hal-03177594>

Submitted on 23 Mar 2021

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.



Distributed under a Creative Commons Attribution 4.0 International License

Université de Droit, d'Économie et des
Sciences
Aix-Marseille III

Faculté des Sciences et Techniques
de Saint-Jérôme

Laboratoire de Valorisation de l'Information

DESS de Gestion des systèmes documentaires et
d'information scientifique et technique

La veille au *Cemagref* : une réalité
quotidienne et un objectif commun.

Rapport de stage

Raymond Schiano
1997

Remerciements

Je tiens à adresser mes remerciements aux personnes suivantes :

- Parina Hassanaly, du Laboratoire de Valorisation de l'Information à la Faculté des Sciences et Techniques de Saint-Jérôme, pour sa patience et la qualité de ses conseils;
- Nicole Delherbe et Karim Haddane pour leurs relectures et leur aide;
- l'ensemble des documentalistes du Cemagref pour leur aide à un moment ou l'autre ;
- et enfin l'ensemble de mes collègues du Cemagref d'Aix-en-Provence pour leur compréhension, leur soutien et leur sympathie.

SOMMAIRE

1. PRÉSENTATION DU CEMAGREF	7
1.1. Les 4 Départements	8
1.1.1. Gestion des territoires	8
1.1.2. Gestion des milieux aquatiques	9
1.1.3. Équipements pour l'eau et l'environnement	10
1.1.4. Équipements agricoles et alimentaires	11
2. INTRODUCTION	12
2.1. Les sources d'information au Cemagref	12
2.2. Le « guide de la veille » :	12
2.3. Le cas des bases de données	12
3. GUIDE DE LA VEILLE AU CEMAGREF	13
3.1. L'accès à l'information : une nécessité vitale dans le monde de la recherche	13
3.1.1. Les enjeux de l'information	13
3.1.2. La veille technologique : la notion	13
3.2. Panorama de l'offre disponible en général	15
3.2.1. Quelques chiffres	15
3.2.2. Panorama des sources d'information	16
3.2.2.1. La presse (les périodiques) :	16
3.2.2.2. Les revues de sommaires	17
3.2.2.3. Les livres (les « ouvrages »)	17
3.2.2.4. Les encyclopédies	17
3.2.2.5. Les thèses	17
3.2.2.6. La littérature grise	18
3.2.2.7. Les brevets	18
3.2.2.8. Les missions et voyages d'études	18
3.2.2.9. Les expositions et salons	18
3.2.2.10. Les comités divers	19
3.2.2.11. Les colloques, congrès, clubs	19
3.2.3. Panorama de la diffusion de l'information	19
3.2.3.1. Les banques de données (et CD-Rom)	19
3.2.3.1.1. Quelques points de vocabulaire	19
3.2.3.1.2. Les différents types	19
3.2.3.2. Internet	20
3.2.3.2.1. Internet, qu'est-ce que c'est	20
3.2.3.2.2. Les services de base	20
3.2.3.2.3. Internet et la recherche d'information	20
3.2.3.3. Les fournisseurs de documents primaires	22
3.2.3.3.1. Le réseau des bibliothèques universitaires et des organismes de recherche publics	22
3.2.3.3.2. Les organismes spécialisés :	22
3.3. De la réflexion naît l'action ou quelques conseils rapides de méthode préalable à toute recherche documentaire	23
3.4. Les sources d'information au Cemagref	24
3.4.1. Les sources internes	24
3.4.1.1. Les autres chercheurs	24
3.4.1.2. Les documentalistes et les centres de documentation	25
3.4.1.3. Les collections de périodiques	25
3.4.1.4. Les fonds documentaires	26
3.4.1.5. La base Cemagref publications	27
3.4.1.6. Les comptes-rendus de missions à l'étranger	27
3.4.2. Les sources externes	27
3.4.2.1. Les bases de données	28
3.4.2.2. Panorama des bases utilisées au Cemagref	28
3.4.2.3. Les Current Contents	29

3.4.2.4. Les serveurs Web	30
3.4.2.5. Le serveur Web du Cemagref	30
3.5. Savoir interroger les banques de données	31
3.5.1. Les étapes de recherche	31
3.5.1.1. Identifier le sujet	31
3.5.1.1.1. Deux techniques préalables	31
3.5.1.1.2. L'entrevue :	31
3.5.1.1.3. Le brainstorming	31
3.5.1.1.4. Comment identifier le sujet?	32
3.5.2. Formulation du sujet	32
3.5.2.1. Construction de la stratégie de recherche : les opérateurs de recherche	33
3.5.2.2. Les opérateurs booléens	33
3.5.2.3. D'autres opérateurs ou procédures de recherche	33
3.5.3. L'exécution de la recherche	34
3.5.4. L'évaluation de l'information trouvée	34
3.5.4.1. Le silence	34
3.5.4.2. Le bruit	34
4. BASES DE DONNÉES : LES PRINCIPAUX MÉTHODES ET OUTILS	35
4.1. Rappel de la problématique	35
4.2. Présentation des principaux outils de recherche d'information	35
4.2.1. Les outils sur papier	36
4.2.2. Les outils en ligne	36
4.2.3. Qu'est-ce qui caractérise la recherche d'information du chercheur ?	36
4.2.4. Diagramme synthétique	37
4.3. Développement des méthodes et outils	38
4.3.1. Les auteurs	38
4.3.1.1. Commande de base de Dialog: SELECT	38
4.3.1.2. Les descriptifs techniques des bases	38
4.3.1.3. Les auteurs et la commande EXPAND	38
4.3.1.5. La commande RANK appliquée aux auteurs	42
4.3.2. Les revues	42
4.3.2.1. Les outils en ligne	42
4.3.2.1.1. Les bases de données de revues	42
4.3.2.1.2. Le Dialog Journal Name Finder	42
4.3.2.2. Champ revues et Commande SELECT	44
4.3.2.3. Champ revues et Commande EXPAND	44
4.3.3. Lieux de congrès	44
4.3.3.1. Les outils en ligne	44
4.3.4. Lieux et noms de labos	45
4.3.4.1. Les outils en ligne	45
4.3.5. Les noms de bases de données	45
4.3.5.1. Les répertoires de bases de données	45
4.3.5.2. La recherche multibases sur serveur	45
4.3.6. Les mots-clé	45
4.3.6.1. La recherche par mot-clé dans les bases de données	45
4.3.6.1.1. L'étape « sémantique »	46
4.3.6.1.1.1. Le dialogue	46
4.3.6.1.1.2. Concepts et mots-clés	46
4.3.6.1.1.3. Les autres paramètres	46
4.3.6.1.2. L'étape « technique »	47
4.3.6.1.2.1. stratégie de base	47
4.3.6.2. La recherche par mots-clé: la commande SELECT de Dialog et ses possibilités	47
4.3.6.2.1. Rappel sur l'index de base	47
4.3.6.2.2. Les troncatures et opérateurs logiques	47
4.3.6.2.2.1. les troncatures	47
4.3.6.2.2.2. les opérateurs de proximité	48
4.3.6.3. Les autres commandes : EXPAND et RANK	51
4.4. ÉTUDE DE CAS	52
4.4.1. Contexte de l'étude	52

4.4.2. L'étude bibliographique	52
4.4.3. La préparation de l'interrogation	52
4.4.4. Première étape : l'identification de bases pertinentes	54
4.4.5. Premier essai sur une seule base	59
4.4.6. L'affinage de la stratégie	64
4.4.7. Comment restreindre le nombre de références?	65
4.4.7.1. Affiner le vocabulaire :	65
4.4.7.2. La commande REMOVE DUPLICATE (RD):	65
4.4.7.3. la restriction :	65
4.4.8. L'interrogation multi-bases	66
4.4.9. Interrogation complémentaire	69
4.4.10. Rappel de la stratégie	69
5. SYNTHÈSE ET PERSPECTIVES	70
5.1. Les autres serveurs	70
5.2. Synthèse des stratégies de base	71
5.2.1. Bilan des critères de pertinence	72
5.2.1.1. limites et intérêt de la méthode	72
5.2.2. L'interrogation multibases : avantages et inconvénients	73
5.2.3. Rappel sur les « Bluesheets »	73
5.2.4. Projet de fiche d'interrogation de bases	73
5.3. Autres éléments de réflexion ou d'actualité	76
5.3.1. L'existence du groupe « fouille de données »	76
5.3.2. La conférence TREC	76
5.3.3. Le problème Questel	76
5.4. La notion de « debriefing » et de retour d'expérience	76
5.4.1. Le debriefing	76
5.4.2. Propositions de mise en commun	79
5.5. Conclusion	80
6. BIBLIOGRAPHIE	82
7. ANNEXES	83
7.1. Annexe 1 : les opérateurs Booléens	83
7.2. Annexe 2 : extrait de la notice EndLink	83
7.3. Annexe 3 : courrier électronique sur le groupe fouille de données	83

1. PRÉSENTATION DU *Cemagref*

Le *Cemagref* est un organisme public à caractère scientifique et technologique (EPST). Créé en 1981, il est sous la double tutelle du Ministère de l'Enseignement Supérieur et de la Recherche et de l'Agriculture et du Ministère de la Pêche et de l'Alimentation.

Son budget a été de 370 M.F. (en 1996) dont 100 M.F. de recettes sur contrats.

Il emploie près de 1000 personnes dont 480 ingénieurs d'origine assez diverses. À ce chiffre viennent s'ajouter une centaine de thésards ainsi que près de 500 stagiaires par an.

Les activités scientifiques du *Cemagref* sont organisées autour de quatre départements. Finalisés et pluridisciplinaires, ils sont au carrefour des champs thématiques et des formes de la production.

Le *Cemagref* mène des recherches finalisées en ingénierie de l'agriculture et de l'environnement. Ses principaux domaines sont :

- les eaux continentales,
- les équipements agricoles et alimentaires,
- la gestion des territoires.

Les options de recherche fixées sont mises en œuvre par une communauté de travail originale qui rassemble plus de mille personnes d'origines diverses et de compétences complémentaires.

Au sein de quatre départements scientifiques, des programmes de recherches sont menés par des équipes d'ingénieurs, de chercheurs et de techniciens réparties sur l'ensemble du territoire français, en dix groupements.

Dans son rôle d'appui à l'action publique, le *Cemagref* met à la disposition d'acteurs de terrain, des avis, des méthodologies, des résultats d'essais, des logiciels, des guides techniques.

Le *Cemagref* produit également des publications scientifiques et techniques, participe et organise des journées d'études, des colloques. Il contribue à l'enseignement supérieur et développe une activité de collaborations internationales importantes.

Sur le plan documentaire, les dix groupements du *Cemagref* comportent 13 fonds documentaires rassemblant plus de 80 000 références. Le logiciel documentaire actuel est Texto Windows de Chemdata, et la mise en commun des références a été faite sur la plateforme WAIS, accessible en interne (zone réservée Extranet) via le serveur Web de l'Établissement.

1.1. Les 4 Départements

1.1.1. Gestion des territoires

La gestion et le développement durable des territoires nécessitent la prise en compte simultanée du fonctionnement des milieux naturels et des activités humaines.

Le département Gestion des Territoires mène des recherches destinées à construire les outils et les méthodes pour la gestion des milieux naturels, des systèmes forestiers et agricoles, et pour le développement local.

Ces recherches prennent en compte les interactions entre les systèmes naturels et sociaux (patrimoine, activités, cadre de vie et risques), et intègrent le rôle des différents acteurs principalement forestiers et agricoles dans la dynamique territoriale.

À l'interface des activités humaines et des milieux naturels

La cible de ses travaux caractérise le département gestion des territoires. Il s'agit en effet d'une part des gestionnaires des milieux forestiers et agricoles, d'autre part des responsables des décisions publiques concernant ces milieux, à savoir l'Union Européenne, les administrations nationales et leurs services déconcentrés, les collectivités territoriales. L'appui technique et l'expertise proposés par ce département scientifique permettent en même temps de valider les méthodes et résultats obtenus, et d'enrichir les futures recherches.

Les travaux tendent vers des réponses pragmatiques en matière de gestion territoriale, mais également des éléments de politique de cette gestion. Les disciplines scientifiques naturalistes et socio-économiques se conjuguent pour aboutir à des modèles transférables.

Les travaux se répartissent au sein de sept programmes:

- gestion des écosystèmes forestiers et pastoraux : étude des relations entre facteurs du milieu, pratiques et production, pour les gestionnaires publics ou privés d'agro-écosystèmes peu intensifs, de forêts ou d'espaces dégradés
- création des peuplements forestiers : conservation des ressources génétiques et production de graines, matériel et techniques de reboisement
- fonctionnement des systèmes forestiers et agro-forestiers : modélisation de la croissance des peuplements forestiers, analyse systémique de certains peuplements (populiculture intensive, forêt paysanne, forêt de montagne)
- paysage et valeurs patrimoniales dans l'aménagement de l'espace rural : étude du paysage dans des milieux divers, évaluation des usages non marchands des milieux naturels, risques d'incendie en forêt et aménagement du territoire
- adaptation des exploitations agricoles : suivi technico-économique d'échantillons d'exploitations agricoles, analyse systémique de leur fonctionnement, typologie, modélisation socio-économique, modes d'utilisation de l'espace rural : analyses des enjeux et des acteurs vis à vis du foncier, de l'échelle de l'exploitation à l'échelle régionale, modélisation spatiale
- territoires ruraux et développement local : analyse systémique à l'échelle d'un territoire des composantes écologiques et socio-économiques, approche par projet, recherche-action.

1.1.2. Gestion des milieux aquatiques

Les unités de recherche du département Gestion des Milieux Aquatiques (GMA) acquièrent et assemblent les connaissances, construisent les méthodes pour aboutir à des recommandations, des outils et des expertises pour la gestion des milieux aquatiques continentaux.

En réponse aux demandes formulées au département GMA, les milieux aquatiques sont considérés dans leurs relations aux activités humaines sous un angle triple : en tant que milieux physiques et milieux vivants, en interaction forte avec les ressources en eau qui les constituent ; en tant que milieux producteurs de ressources vivantes exploitées ; enfin en tant que milieux récepteurs des rejets bruts ou épurés.

Des cultures et des compétences pour intégrer recherche et action

Les unités de recherche s'efforcent de concevoir des méthodes, des recommandations et des outils pour les gestionnaires, qui puissent concilier la préservation du patrimoine, l'utilisation et la mise en valeur de la ressource.

Les démarches employées associent étroitement les observations et mesures sur le terrain, sur des sites naturels ou en stations expérimentales, les travaux de laboratoires et la modélisation physique ou numérique des systèmes étudiés.

Les travaux sont organisés en cinq programmes dont les interactions vont en se développant :

- Le programme sur le fonctionnement hydrologique des bassins et des réseaux hydrographiques vise à construire les méthodes et outils d'une gestion équilibrée, quantitative et qualitative de la ressource en eau.
- Le programme sur les déterminants physiques du fonctionnement écologique des hydrosystèmes d'eau courante se consacre à l'aménagement et la gestion équilibrée du milieu aquatique.
- Le programme sur les altérations physico-chimiques et biologiques des écosystèmes aquatiques construit des méthodes pour fixer les contraintes imposées aux rejets et aux activités, tenant compte des potentialités des milieux récepteurs.
- Dans le programme sur les ressources aquatiques vivantes exploitées sont développés les méthodes et outils de leur gestion. Il s'agit de soutenir des politiques publiques sur les peuplements de poissons.
- Enfin dans le programme épuration des eaux, les nouveaux procédés d'épuration des eaux sont évalués et les méthodes d'optimisation des performances sont développées.

1.1.3. Équipements pour l'eau et l'environnement

Le département Équipements pour l'Eau et l'Environnement (EEE) du *Cemagref* mène des recherches en ingénierie sur les équipements destinés à une ressource naturelle essentielle, l'eau, ainsi qu'à l'environnement. Ses finalités sont l'usage agricole de l'eau, la maîtrise de certains risques liés à l'eau, des services publics tels que l'adduction d'eau potable et les déchets municipaux.

L'environnement (l'impact des équipements), la sécurité (celle des personnes mais également des biens et des fonctionnements), les performances techniques et économiques sont des éléments importants de la maîtrise des usages agricoles et domestiques de l'eau, et des risques qui lui sont associés.

Des recherches consacrées aux outils et méthodes de diagnostic et de gestion

Les recherches du département EEE conduisent au développement d'outils pour la conception et l'exploitation d'équipements ou d'aménagements, en particulier ceux qui contribuent à l'utilisation raisonnée de l'eau pour l'agriculture. Le système " sol-eau-ouvrage " est au point de départ des préoccupations, qui comprennent également l'évaluation micro-économique des périmètres irrigués et drainés.

Des outils et méthodes sont également conçus pour diagnostiquer les risques liés à l'eau, en particulier la sécheresse, les érosions, les ruptures de barrages et les avalanches. Pour tous ces risques, l'objectif est de proposer des méthodes de surveillance et de prévention, ou des ouvrages de protection.

Enfin, les travaux du *Cemagref* sur les services publics améliorent l'efficacité techniques des réseaux d'eau potable et des installations de déchets municipaux, ainsi que les procédures de gestion de ces services.

Les recherches sont réparties dans six programmes nationaux :

- ingénierie et prévention de risques naturels en montagne,
- ingénierie et sécurité des barrages et des ouvrages hydrauliques,
- maîtrise et gestion des déchets,
- techniques et gestion de l'irrigation et du drainage,
- techniques et gestion d'équipements publics pour l'eau et les déchets,
- ingénierie des interactions eau-matériaux-ouvrages.

Le département EEE produit des connaissances et expertises, des essais et des normes, des logiciels, des guides techniques et méthodes, des formations. La création de logiciels de conception, d'optimisation ou d'aide à la gestion constitue un point fort du département.

1.1.4. Équipements agricoles et alimentaires

Les équipements facteurs de production dans la chaîne alimentaire

Les recherches du département sont consacrées aux agro-équipements pour la production primaire, au contrôle des process et à l'instrumentation pour la transformation secondaire, ainsi qu'aux systèmes de stockage et de transport (frigorifique) pour la logistique et la distribution finale. Ces équipements mobilisent des démarches scientifiques et technologiques communes.

Dans le secteur de la production agricole

Les recherches visent à doter les agro-équipements de technologies modernes déjà introduites dans d'autres secteurs industriels : électronique et informatique embarquées, robotique et vision artificielle, mécanique avancée, systèmes d'information. De nouvelles machines et de nouveaux systèmes de pensée émergent, débouchant sur de nouveaux outils de gestion et d'aide à la décision.

Dans le secteur des industries agro-alimentaires

Les recherches privilégient les pilotes de recherches et les démarches par modélisation. Elles portent sur des systèmes spécifiques tels les capteurs, la métrologie industrielle, les systèmes frigorifiques, mais également le traitement du signal, les approches multi-sensorielles, la mécanique des fluides ou l'intelligence artificielle.

Le département incorpore les avancées de la biologie végétale et animale, les nouveaux itinéraires techniques, et les besoins du marché en innovations. Des partenariats sont recherchés en agro-alimentaire pour les analyses (biologiques, biochimiques etc.), le génie des procédés alimentaires, les analyses sensorielles ou encore avec les professionnels de la chaîne du froid.

Cinq programmes de recherche, deux modes de valorisation

Les recherches et développements du département sont conduites au sein de 7 programmes :

- Mécafutur pour la conception de machines intelligentes pour l'agriculture, la sylviculture et l'entretien des espaces naturels,
- Novaprocess pour la conception de systèmes d'équipements complexes pour l'agro-industrie,
- Agripropre pour la maîtrise de l'impact des agro-équipements sur l'environnement, Sensoral pour les mesures et le contrôle rapide de la qualité alimentaire,
- Frigéqual pour la maîtrise des procédés frigorifiques et de la chaîne du froid alimentaire,
- Certimach pour les essais et expertises pour la certification des équipements agricoles et forestiers, Mecaplus pour les prestations d'ingénierie et d'essais de terrain pour les organismes interprofessionnels et les PME/PMI.

Les recherches du département doivent conduire également à la mise au point d'outils d'aide à la décision pour le choix des investissements, et d'outils d'analyse de l'opportunité et d'évaluation des équipements innovants. Finalisées, elles vont jusqu'à la mise à l'épreuve expérimentale, avec les instituts techniques et les industriels, par la réalisation de démonstrateurs développés en commun avec les industriels et utilisateurs. Ses experts participent à des actions de normalisation et d'essais en matière de sécurité, de protection de l'environnement et des opérateurs etc... La valorisation du savoir-faire acquis passe par son transfert rapide au sein du tissu industriel.

2. INTRODUCTION

2.1. Les sources d'information au *Cemagref*

À la suite d'une enquête sur la fonction documentaire au *Cemagref*, on a conclu à la nécessité d'améliorer :

- d'une part la connaissance générale des sources d'information disponibles;
- d'autre part la connaissance des bases de données commerciales en ligne, peu exploitées.

Le document qui suit se composera donc de deux grandes parties :

- un « Guide de la veille » à l'intention des ingénieurs-chercheurs du *Cemagref*;
- un travail sur l'utilisation des bases de données commerciales en ligne, recensant quelques méthodes et outils.

2.2. Le « guide de la veille » :

Le but principal de ce document est d'apporter à un public d'ingénieurs-chercheurs au *Cemagref* quelques connaissances simples en matière d'information scientifique et technique :

- un panorama de l'offre disponible en général,
- une présentation de ce qui est disponible au sein du *Cemagref*,
- et quelques éléments de méthode à propos de la recherche documentaire sur bases de données.

2.3. Le cas des bases de données

Cette partie porte d'abord sur l'examen des principaux outils et moyens « d'extraire » l'information de ces bases. Puis, au travers d'une étude de cas, c'est-à-dire une interrogation réelle répondant à un besoin, on décrit la méthode qui a permis d'obtenir des résultats, puis on tente d'en extraire des éléments opérationnels, applicables à un plus grand nombre de cas.

3. GUIDE DE LA VEILLE AU CEMAGREF

3.1. L'accès à l'information : une nécessité vitale dans le monde de la recherche

3.1.1. Les enjeux de l'information

Affirmer que celui qui détient l'information détient le pouvoir semble être devenu un lieu commun. Et pourtant...

L'information est diverse, multiforme, multilingue; elle est dépendante d'un environnement donné, culturel, scientifique, dans lequel elle s'échange sur différents supports ou dans un cadre institutionnel.

L'information est à la fois quelque chose que l'on utilise souvent, que l'on traite parfois mais aussi que l'on produit.

Cela suppose que chacun prenne conscience qu'il est forcément impliqué dans la circulation de l'information et qu'il doit la maîtriser s'il ne veut pas qu'elle lui échappe.

Ainsi, un chercheur abordant par exemple une nouvelle thématique devra :

- faire le point sur le sujet, connaître les réalisations, les hypothèses posées au départ, celles déjà vérifiées et les autres, ...
- s'informer sur les autres chercheurs et équipes de recherche travaillant sur le même sujet,
- identifier les sujets annexes susceptibles de l'intéresser.

En effet, l'information fournie par les autres est importante non seulement pour voir ce qui est fait sur une même thématique, mais aussi sur les thématiques voisines pour amener de nouvelles idées ou éléments de méthodologie.

3.1.2. La veille technologique : la notion

Cette notion, également appelée intelligence économique, est apparue en France au cours des années 80. Par méthodes de veille technologique, il faut comprendre : tout ce qui permet à un établissement d'essayer de déterminer les secteurs d'où viendront les innovations majeures de son métier, tant pour les procédés que pour les produits. Cela inclut notamment le repérage des principaux acteurs et de leurs stratégies.

Pourquoi faire ?

Le concept s'est d'abord beaucoup développé dans les entreprises, et le secteur recherche et développement et marketing. Le but est de repérer les possibilités d'évolution, adapter l'entreprise, en un mot orienter **l'innovation**.

Ces notions peuvent-elles s'appliquer au Cemagref ?

Si un EPST comme le Cemagref n'a pas le même positionnement économique qu'une entreprise, la valeur de l'information y est tout aussi élevée. Les notions décrites ci-dessus s'appliquent parfaitement à l'environnement scientifique et technique.

Ceci peut s'illustrer par l'exemple d'un thésard arrivant au Cemagref. Une fois son sujet défini dans les grandes lignes, il va se lancer dans une recherche d'informations (la phase "bibliographie"). Cette phase vise à acquérir l'ensemble des connaissances disponibles dans la communauté scientifique internationale, et permet d'obtenir des repérages sur deux axes :

- le sujet : vérifier que celui-ci est bien posé, qu'il s'inscrit dans la bonne thématique de recherches
- les méthodes : l'investigation de la thématique de recherches permet de dégager des méthodes utilisables; si aucune ne se dégage clairement, l'examen de thématiques proches servira à identifier des méthodes extrapolables.

3.2. Panorama de l'offre disponible en général

Ce chapitre décrit l'offre en matière d'information scientifique et technique telle qu'elle est généralement disponible à un public d'ingénieurs, chercheurs et universitaires.

3.2.1. Quelques chiffres

Le développement spectaculaire de la science et de l'innovation technologique peut s'illustrer par une image forte, sinon totalement fidèle. Selon la National Education Association américaine, "il a fallu attendre 1750 pour que le savoir de l'homme de l'époque du Christ puisse seulement être doublé. Une nouvelle multiplication par deux s'est achevée 150 ans plus tard, en 1900... La quatrième multiplication de ce genre, pour l'ensemble du savoir scientifique, s'est déroulée dans la seule décennie qui suit 1950. D'une autre manière, on peut dire que la connaissance technologique a été multipliée par dix tous les 50 ans depuis plus de 2800 ans. En 1950, on comptait 1 million de chercheurs et d'ingénieurs dans le monde. En 1900, ils étaient au nombre de 100 000, en 1850 10 000 et en 1800 1 000." Ce chiffre dépasse sans doute largement les dix millions aujourd'hui.

En supposant que la production est proportionnelle au nombre de chercheurs, on imagine aisément la progression de la quantité d'informations et de résultats disponibles.

Quelques autres chiffres permettent de mesurer l'ampleur du phénomène :

- De 10 000 titres par an au début du XX^{ème} siècle, la production mondiale de périodiques était de 170 000 en 1971, et proche du million en 1995
- 5646 banques de données sont accessibles en ligne début 1995
- Minitel : parmi les 24600 services proposés fin 1994, 235 correspondent à des banques de données pour l'entreprise et à des services scientifiques et techniques (en France principalement).
- Pour les banques de données sur CD-Rom, le catalogue mondial comptait début 1994 5379 titres (produits à contenu scientifique et culturel, source TFPL Publishing). Ce chiffre est proche de 8000 début 1995 .

- Quelques dates parlantes :
 - années 60 : premiers fonds documentaires informatisés
 - 1970 : création du Centre de Documentation Scientifique et Technique (CDST) et du Centre de Documentation en Sciences Humaines (CDSH), devenus l'INIST
 - 1972 : les premiers réseaux
 - 1979 : le premier serveur français
 - 1981 : naissance du Minitel et du micro-ordinateur
 - 1986 : apparition commerciale du CD-Rom
 - 1990 : maillage mondial des réseaux électroniques

On constate donc non seulement une explosion de l'offre mondiale, mais aussi une montée en puissance énorme des moyens d'y accéder.

3.2.2. Panorama des sources d'information

Nous allons présenter un peu plus en détail les principales sources d'information scientifique et technique et leurs caractéristiques.

3.2.2.1. La presse (les périodiques) :

- Les revues scientifiques présentent généralement la structure suivante :
 - une périodicité hebdomadaire, mensuelle, bimestrielle, semestrielle ou annuelle
 - une page de couverture identique pour chaque livraison sur laquelle figurent les mentions de titres, d'année, de volume, ...
 - la liste du comité de rédaction, les conditions d'abonnements, les instructions aux auteurs, un sommaire...

Les articles eux-mêmes ont des présentations variées mais on y retrouve toujours les éléments suivants :

- le titre,
 - les noms et prénoms des auteurs suivis de leur affiliation, c'est-à-dire le nom et adresse de l'organisme où ils travaillent,
 - un résumé, le plus souvent en anglais, et/ou un résumé dans la langue de l'article,
 - une liste de références bibliographiques.
- Un point spécifique de ces revues est le système de validation par des pairs : un comité de lecture composé de spécialistes reconnus permet une confrontation des résultats décrits.
 - Cette presse reste le principal moyen de diffusion des résultats de la recherche. De plus, comme elle est très "référéncée", elle constitue l'information la plus facile d'accès.

L'ISI et le système de cotation des périodiques

L'Institute for Scientific Information (ISI), a été créé par Eugène Garfield en 1958, dans le but de donner accès aux chercheurs aux sommaires de revues scientifiques. Il met maintenant à leur disposition un certain nombre de banques de données relatives aux publications, et se prêtant particulièrement bien à l'analyse statistique.

L'ISI traite, pour l'ensemble de ses produits, 8000 périodiques à comité de lecture (peer-reviewed journals).

Parmi l'importante gamme de produits documentaires proposés par l'ISI, les plus connus sont le Science Citation Index (SCI), banque constituée par les citations des articles; ceux-ci constituent eux-mêmes, avec leurs résumés, la base des Current Contents.

Le SCI

Présenté sous forme d'index, le SCI dépouille près de 5000 périodiques du monde entier (avec une prédominance des revues de langue anglaise). Il permet aux chercheurs de retrouver des références bibliographiques d'articles sur un sujet donné à partir de différents critères de recherche : nom d'un auteur, affiliation des auteurs, mots du titre, articles cités.

Le SCI Journal Citation Reports

À partir de ces dépouillements, le SCI élabore chaque année le Journal Citation Reports (JCR), qui permet d'effectuer un classement des périodiques dépouillés, selon divers critères, et plus particulièrement selon le nombre de citations dont ils ont fait l'objet dans d'autres périodiques.

Ces classements de périodiques permettent d'établir une sorte de "hit-parade" des périodiques les mieux considérés, c'est-à-dire les plus cités, dans le domaine qui les intéresse.

Bien sûr un tel classement n'est pas sans effet ni sur la politique éditoriale des revues, ni, surtout, sur le choix des chercheurs quant aux revues auxquelles ils vont, de préférence, adresser leurs articles pour demande de publication.

3.2.2.2. Les revues de sommaires

Il s'agit d'un produit largement répandu : la plupart des centres de documentation produisent désormais ce type de produits pour informer leurs utilisateurs. Elles peuvent être sur support papier ou informatisé.

Le produit informatisé le plus connu sont les Current Contents publiés par l'ISI. Il s'agit d'une bibliographie scientifique reçue périodiquement sur disquettes (ou sur serveur), destinée à être consultée régulièrement à l'aide de profils de recherche (voir plus loin le paragraphe qui leur est consacré).

3.2.2.3. Les livres (les « ouvrages »)

Ce sont souvent des synthèses : l'information a été traitée par l'auteur. Certains ouvrages dits "de référence" sont réputés contenir la base de certaines connaissances.

3.2.2.4. Les encyclopédies

Elles ne doivent pas être négligées. Elles peuvent constituer la source idéale pour débroussailler un sujet et son environnement ou préciser des aspects de vocabulaire.

3.2.2.5. Les thèses

Elles présentent des caractéristiques très semblables aux ouvrages.

Avantages : ils résident dans la synthèse bibliographique et éventuellement dans les résultats.

Inconvénients : il s'agit d'un exercice universitaire à valeur parfois inégale, la lecture en est longue et rend compliqué m'accès à l'innovation apportée..

L'accès au document est toujours possible pour la France, via le réseau du prêt inter-bibliothèques universitaires.

Téléthèse est le catalogue des thèses soutenues devant les universités françaises. Il référence environ 300 000 thèses. Il est consultable :

- Minitel 36 15 code SUNK*THE
- Minitel 36 16 code SUNIST*THE
- CD-Rom DocThèses dans les bibliothèques universitaires (mise à jour semestrielle)

3.2.2.6. La littérature grise

On désigne sous l'appellation de "littérature grise" l'ensemble des rapports, travaux universitaires, communications ou actes de congrès produits et parfois diffusés par les entreprises et organismes mais qui ne font pas l'objet d'une édition.

On peut donc ranger dans cette catégorie :

- les rapports de recherche ou d'étude produits par des organismes publics ou privés,
- les pré-publications ou preprints,
- les comptes-rendus de congrès non édités,
- les documents publicitaires et notices de constructeurs.

3.2.2.7. Les brevets

Tout inventeur d'un nouveau procédé ou technique doit protéger son invention. Le dépôt d'un brevet (en anglais : patent) protège les droits d'exploitation pour une durée et une zone géographique déterminées. Le dépositaire d'une demande doit décrire son invention dans une publication : le brevet.

Dans le domaine de la veille technologique, la documentation contenue dans les brevets constitue une des sources les plus riches. Le contenu informatif du brevet est exploité pour les applications suivantes :

- surveillance globale des secteurs techniques
- surveillance technique précise, sectorielle
- surveillance de l'activité de la concurrence
- détermination du domaine protégé.

3.2.2.8. Les missions et voyages d'études

Bien qu'onéreuses, ce sont des sources riches en information. La valeur ajoutée de ces missions, ce sont les comptes-rendus faits par les participants et l'intérêt qu'ils présentent pour les autres.

3.2.2.9. Les expositions et salons

Leur intérêt principal réside dans le regroupement professionnel qu'ils favorisent : ils coûtent donc moins chers que les missions en termes de contacts. Il faut simplement s'adapter

à la brièveté et à la densité de choses à voir. Penser à ramener un catalogue d'exposition : ces catalogues peuvent s'avérer une mine de renseignements.

3.2.2.10. Les comités divers

Tous les comités où se rencontrent des gens d'organisations différentes (syndicats interprofessionnels, comités de lecture ou de normalisation, commissions ad hoc,...) sont des sources d'informations.

3.2.2.11. Les colloques, congrès, clubs

Leur importance est primordiale; la plus grande partie de leur intérêt réside souvent autant dans les communications que dans les contacts que l'on y noue et dans les informations qui circulent à ce moment. En ce qui concerne les colloques, de nombreuses communications sont désormais éditées; lorsqu'il y a comité de lecture, ils s'apparentent alors aux revues

3.2.3. Panorama de la diffusion de l'information

3.2.3.1. Les banques de données (et CD-Rom)

3.2.3.1.1. Quelques points de vocabulaire

- les banques de données sont un ensemble de données relatif à un domaine défini de connaissances et organisé pour être offert aux consultations d'utilisateurs. On utilisera indifféremment dans ces pages les termes "bases" et "banques" de données car la différence entre des deux notions tend à s'estomper. Ces bases sont éditées par un éditeur ou producteur, qui confie généralement leur mise en accès à un serveur.

- le serveur :

Organisme exploitant un système informatique permettant à un demandeur de consulter et d'utiliser directement une ou plusieurs banques de données fournies par différents producteurs. Il assure la commercialisation de ces produits et développe le plus souvent un logiciel commun pour leur interrogation.

- Avantages :

- .les serveurs sont normalement très exhaustif

- Inconvénients :

- . le laps de temps d'édition est de l'ordre de 6 mois à un ans. Aux délais de publication s'ajoutent ceux de l'introduction dans la banque.

- . leur interrogation nécessite un certain professionnalisme

- . il est parfois peu facile d'obtenir les documents originaux.

3.2.3.1.2. Les différents types

Suivant la nature de l'information enregistrée, on distingue :

- les **banques bibliographiques** contenant des références bibliographiques parfois accompagnées de résumés,

- les **banques répertoires** contenant des caractéristiques d'entreprises ou des biographies,

- les **banques textuelles** offrant le texte intégral des documents (texte juridique, presse quotidienne,...)

- les **banques de données factuelles** contenant des données autres que du texte (données économiques, formules chimiques,...).

Cette liste n'est pas exhaustive, il existe notamment des banques de données multimédia indexant du texte, du son et de l'image.

Enfin, on voit souvent les termes de banque de données "**ASCII**" et "**vidéotex**" : les premières sont composées de texte au standard ASCII; c'est le cas notamment des banques bibliographiques. Les secondes sont des banques accessibles par Minitel.

3.2.3.2. Internet

3.2.3.2.1. *Internet, qu'est-ce que c'est*

- Internet est un terme générique qui désigne la multitude des réseaux utilisant un certain type de protocole (TCP/IP) dans le monde.

- En France, le réseau Renater (Réseau national de télécommunication pour la technologie, l'enseignement et la recherche) assure la liaison avec Internet.

3.2.3.2.2. *Les services de base*

- **le courrier électronique (mail)** : il suffit de connaître l'adresse électronique de son correspondant pour lui faire parvenir un message.

- **le transfert de fichier (FTP)**, File Transfer Protocol : permet d'expédier et d'aller chercher des fichiers. Dans ce second cas, il faut aussi connaître l'adresse du site où se trouve le fichier ainsi que son nom.

- **Telnet** : ouverture d'une session de travail sur une machine distante

- **WWW** ou les pages **Web** : ces pages représentent plus de 50% du trafic. Elles proposent graphiques et photos, et des liens hypertexte permettant une navigation aisée de page à page ou de site à site (cf plus bas).

3.2.3.2.3. *Internet et la recherche d'information*

La prolifération des informations sur Internet ainsi que leur dispersion ont très vite fait émerger la nécessité d'outils spécifiques pour structurer les données et surtout pour les retrouver et savoir les chercher.

En nous limitant au domaine de l'information scientifique et technique, voici quelques produits et services souvent cités :

- **WWW** (World Wide Web): système hypermédia de gestion de l'information. Il devient actuellement l'outil standard sur Internet. Hypermédia signifie que l'on clique sur un

mot ou une zone soulignée, et cela renvoie soit à une autre page soit à un autre site. On accède au serveur WWW (ou Web) par des modules clients tels que Netscape ou Mosaic. Une ressource comporte une "adresse" propre ou URL (Uniform Resource Location); voici l'URL du serveur Web du Cemagref : <http://www.cemagref.fr>

- les **forums** de discussion: grâce au courrier électronique, des milliers de listes de discussion sont disponibles sur toutes sortes de sujets.

- les **news** constituent un autre système pour les forums de discussion. On compte actuellement 5000 groupes de news. Le fonctionnement est différent de celui des forums : l'utilisateur consulte les messages qui l'intéressent en accédant à une machine serveur de news. Il ne reçoit donc pas systématiquement les messages, ce qui permet d'opérer une sélection.

- les **lettres d'information** : on en dénombre près de 400 sur Internet.

Elles constituent un vecteur de communication important entre chercheurs pour connaître le calendrier des futures manifestations scientifiques, les appels à communication, les résumés et commentaires de certains travaux.

- les **revues électroniques** : le réseau devient une alternative pour diffuser rapidement les résultats de la recherche. On recense plus de 300 revues électroniques début 1995.

- les **catalogues de bibliothèques** : de nombreux catalogues sont désormais interrogeables, dont ceux de la bibliothèque du Congrès, du Cern,... Quelques catalogues français sont recensés sur :

<http://www.grenet.fr/anteserveur/france.html>

- **Rechercher l'information à l'aide de "moteurs de recherche**

L'utilisation de moteurs de recherche

Les pages Web en tant que telles contiennent une masse énorme d'information; les moteurs de recherche permettent de "naviguer", avec la limite qu'il faut savoir ce que l'on cherche. Ce sont des programmes informatiques qui se chargent de la recherche. Ils répondent à une requête exprimée généralement à l'aide d'opérateurs booléens (ET, OU, SAUF). Ils travaillent à partir d'autres recensements que ceux des serveurs; les stratégies de recherche peuvent être différentes. Ils ne fournissent donc pas toujours la même réponse aux questions posées. Il peut se révéler utile d'en utiliser plusieurs. Le serveur Web du Cemagref cite et permet d'accéder à un certain nombre d'entre eux, connus pour leurs qualités respectives.

3.2.3.3. Les fournisseurs de documents primaires

Les moyens de se procurer les documents primaires sont multiples. Deux grands types d'organismes assurent ce service : les bibliothèques universitaires organisées en réseau, et les organismes spécialisés.

3.2.3.3.1. Le réseau des bibliothèques universitaires et des organismes de recherche publics

- Un étudiant consulte d'abord le catalogue de sa bibliothèque, et si celle-ci ne possède pas le document elle le demande en prêt inter-bibliothèques : un système de messagerie électronique relie 200 bibliothèques universitaires (BU) et 100 autres établissements.

- Les CADIST (Centres d'acquisition et de diffusion de l'information scientifique et technique), au nombre de 26, renforcent ce réseau : ce sont des pôles ressources dont la mission est de fournir rapidement les documents en prêt.
- Nous avons déjà mentionné Téléthèses, le catalogue des thèses soutenues en France. Deux autres catalogues complètent le dispositif :
- - le CCN (Catalogue collectif national des publications en série) : il recense les collections de périodiques; il propose 672 000 titres dont 262 000 sont localisés dans les bibliothèques adhérentes.
- - le Pancatalogue : catalogue collectif des ouvrages français et étrangers acquis par les BU et établissements relevant du ministère chargé de l'enseignement supérieur. En 1995, il recense 1 million 900 000 titres localisés dans 111 bibliothèques.

3.2.3.3.2. Les organismes spécialisés :

- les **libraires et diffuseurs** : certains se sont spécialisés dans l'obtention de documents scientifiques et techniques, comme Lavoisier ou Dawson.
- les **banques de données** elles-mêmes offrent de plus en plus souvent
 - . du texte intégral
 - . des données factuelles directement exploitables
 - . la possibilité de commander des documents en ligne à partir de références bibliographiques sélectionnées lors de l'interrogation d'un code client.
- la **Bibliothèque Nationale de France (BNF)**

La Bibliothèque nationale de France est une bibliothèque de conservation et de recherche. Elle accueille les chercheurs ou étudiants dont les recherches nécessitent des documents introuvables ailleurs. La consultation de l'ensemble des documents de la BNF n'est possible qu'après l'obtention d'un titre d'accès.

La BNF élabore de nombreux produits documentaires. Elle comprend l'Agence bibliographique nationale dont la mission est de collecter et de signaler la production éditoriale française reçue par dépôt légal. Elle produit notamment la base BN-OPALE (deux millions de références), laquelle est consultable à partir du serveur Web de la Bibliothèque nationale de France (<http://www.bnf.fr>).

- **l'Inist**

<p>L'Inist (Institut de l'information scientifique et technique)</p>
--

<p>L'Inist a pour mission de collecter, traiter et diffuser les résultats de la recherche scientifique et technique internationale. C'est le premier centre français de fourniture de copies de documents scientifiques et techniques. En pratique, les demandes de copies de document nécessitent l'ouverture préalable d'un compte de dépôt et l'attribution.</p>

Ce bref panorama de l'offre mondiale démontre quelque peu son immense diversité pour tous les types d'information.

Le présent guide se propose, dans les pages qui suivent, de fournir quelques références supplémentaires sur certains de ces médias accessibles au Cemagref. Nous décrirons notamment des ressources internes comme les fonds documentaires et les collections de périodiques, des ressources comme les Current Contents, et surtout nous présenterons quelques éléments de méthode simples pour l'accès aux bases de données.

3.3. De la réflexion naît l'action ou quelques conseils rapides de méthode préalable à toute recherche documentaire

Avant de se lancer dans une recherche d'information, il importe de se livrer à une réflexion sur la manière de la faire afin d'éviter les pertes de temps et d'argent :

- se pose-t-on la bonne question?
- a-t-on choisi les bonnes sources?

Par exemple, une équipe souhaite lancer une recherche sur l'anguille :

- il faut d'abord définir le sujet : à quoi va servir l'information?

Ex : étudier les migrations, pour savoir si l'aquaculture est possible

- à partir de là, poser les questions et surtout les expliciter :

Ex : que sait-on sur la migration?

Si on sait "tout", on a les réponses sur la migration;

sinon, on a des hypothèses sur la migration, on commence à voir ce qu'on ne sait pas, et donc aussi ce qu'on doit chercher

- jusqu'où doit-on aller?

Ex : ici, en fonction des premières réponses, il faut décider si on doit continuer à chercher sur la migration (en recherchant l'exhaustivité si c'est nécessaire) ou si on peut passer à la seconde partie de la question, l'aquaculture. Cela implique notamment que l'objectif soit clairement défini.

- le choix des sources

- a priori on va examiner les revues sur l'eau, l'aquaculture et la pêche. L'anguille étant un poisson pêché, on devrait par exemple trouver des articles dans le « Bulletin français de pêche et de pisciculture »

- parallèlement, on va aussi poser la question aux autres chercheurs dans ce domaine, à l'intérieur comme à l'extérieur du Cemagref

- on va ensuite interroger les Current Contents et y enregistrer un profil de recherche sur ce thème

- et bien sûr on pourra interroger les bases de données.

3.4. Les sources d'information au Cemagref

Après avoir fait un panorama général des principales sources d'information scientifique et technique, nous allons présenter les sources plus spécifiques au Cemagref.

3.4.1. Les sources internes

3.4.1.1. Les autres chercheurs

C'est dans son entourage professionnel immédiat que le nouveau chercheur va découvrir une part importante de l'information utile à son activité.

Certains chercheurs ou équipes alimentent des bases de données constituées de leurs éléments bibliographiques propres, lesquelles s'avèrent souvent d'une grande richesse.

L'accès à ce type d'information se fait souvent par bouche à oreille, ou par le biais d'une animation scientifique spécifique.

Voici la liste des groupes et ateliers "transversaux" au Cemagref :

- analyse de données
 - réseau des numériciens
 - machines intelligentes et contrôle de la mobilité
 - écologie terrestre
 - qualité, essais et analyse
 - biochimie et microbiologie
 - coordination des activités des labos du secteur ouvrages hydrauliques du Cemagref
 - mesures physiques dans les sols et normalisation
 - télédétection et SIG
 - économie
- Une précision : ces groupes ne sont pas homogènes dans la mesure où ils fonctionnent différemment, ont des objectifs variés et ne sont pas dans le même stade de maturité...

3.4.1.2. Les documentalistes et les centres de documentation

La documentation au Cemagref se compose avant tout de personnes avec des compétences en matière d'information (les documentalistes), dont le savoir-faire est destiné à rendre un service, et de lieux (les centres), organisés pour accueillir. Une à deux personnes par groupement sont à la disposition de l'ensemble des chercheurs d'un site.

Les documentalistes du Cemagref sont structurés dans un réseau dans lequel circule l'information.

Les missions des documentalistes ont été clairement définies par l'Établissement :

a) Satisfaire prioritairement les besoins des ingénieurs et des chercheurs de l'établissement, ce qui suppose :

- de mettre à leur disposition des bases internes;
- de contribuer à l'activité de veille documentaire réalisée par les ingénieurs-chercheurs en s'appuyant non seulement sur les bases internes, mais aussi sur des ressources externes;

- de les aider à utiliser efficacement et dans un but documentaire les outils qu'ils possèdent déjà, en particulier Internet;

- de les aider à se procurer les documents primaires dont ils ont besoin;

- de leur apporter conseil et assistance pour la gestion de leur documentation personnelle - notamment les tirés à part - et de leurs bibliographies.

b) Contribuer à la valorisation des publications scientifiques et techniques du Cemagref en alimentant le dispositif de dépôt/valorisation des publications,

c) Répondre de façon sélective à une demande externe, dans la mesure où celle-ci génère une plus-value pour l'établissement.

Les documentalistes, pour satisfaire les demandes, disposent d'un certain nombre de moyens dont notamment :

- des collections de périodiques

- des fonds documentaires

- des abonnements à des bases de données.

3.4.1.3. Les collections de périodiques

L'Établissement dans son ensemble est abonné à près de 1800 titres de périodiques scientifiques et techniques.

	Nombre de titres de périodiques reçus en 1995
Aix	187
Antony	172
Bordeaux	228
Clermont Riom	110
Clermont Montoldre	60
Grenoble	203
Direction Générale	200
Lyon	165
Montpellier	243
Nogent	103
Rennes	104
Téledétection	24
Total	1799

Dans la plupart des groupements ils sont consultables en libre accès dans une salle de lecture; les sommaires circulent habituellement dans les divisions concernées. La Dicova réalise à intervalles réguliers une mise à jour du catalogue de tous les périodiques reçus au Cemagref.

Ce catalogue figure également sur le serveur Web.

On peut donc, via les documentalistes, se procurer rapidement (et gratuitement) tout article d'un périodique reçu par un autre groupement.

3.4.1.4. Les fonds documentaires

Ils sont constitués en grande partie d'ouvrages, complétés par des études rapports, congrès,...

L'accès à ces fonds documentaires se fait le plus souvent par l'intermédiaire des documentalistes qui les alimentent et qui en ont une bonne connaissance.

Ces fonds sont informatisés. Ils sont gérés selon la même structure sur tous les sites, à l'aide du logiciel documentaire Texto Windows.

La politique d'achat des abonnements et ouvrages est spécifique à chaque groupement et dans la pratique à chaque équipe.

Fonds documentaires du Cemagref

	Nombre total de références des base de données en décembre 1995
Aix Groupement	12075
Aix Forêt	4700
Antony	13617
Bordeaux	11667
Clermont Riom	5320
Clermont Montoldre	1692
Grenoble	5885
Direction Générale	2295
Lyon	5500
Montpellier	3175
Nogent	2121
Rennes	3691
Téledétection	1117
Total	80230

Les bases de données peuvent donc être consultées en local :

- soit directement à partir d'un poste adéquat,
- soit dans les Centres de Documentation par les documentalistes ou avec leur appui.

L'ensemble des bases de données de l'Établissement sont également consultables sur le serveur Web du Cemagref (voir plus loin).

3.4.1.5. La base Cemagref publications

Elle est gérée par le service IST et disponible sur le serveur Web du Cemagref. Elle regroupe l'ensemble des publications de nature scientifique ou technique rédigées ou co-rédigées par des ingénieurs - chercheurs du Cemagref, y compris la littérature grise.

L'accès au document se fait sur simple demande auprès du service IST.

3.4.1.6. Les comptes-rendus de missions à l'étranger

On peut vérifier auprès du SREI (Service des Relations Européennes et Internationales) s'il existe une mission sur un pays ou un thème donné. Dans l'affirmative et sur simple demande, ce service se charge de communiquer ces comptes-rendus de missions à l'étranger.

3.4.2. Les sources externes

Les sources externes que nous allons décrire sont celles qui sont utilisables au Cemagref, ou depuis le Cemagref, c'est-à-dire des médias autorisant des recherches d'informations externes. Ce sont essentiellement les bases de données, les Current Contents et les serveurs Web.

3.4.2.1. Les bases de données

Potentiellement, toute base est accessible au Cemagref. Dans la pratique, les documentations des groupements prennent des abonnements à un ou plusieurs serveurs.

De plus, un certain nombre de serveurs sont accessibles directement via Internet, et donc sans abonnement. Une étude réalisée en 1995 par la Dicova, également présente dans tous les centres, recense les 24 bases interrogées au Cemagref (le tableau ci-dessous énumère ces bases). Pour chacune d'entre elles figurent des informations relatives au contenu, au producteur, aux moyens d'accès par groupement. Sont également cités les centres de documentation ayant une expérience de leur utilisation.

L'une des conclusions de cette étude est que les bases de données susceptibles d'intéresser les chercheurs du Cemagref sont mal connues de ces derniers.

3.4.2.2. Panorama des bases utilisées au Cemagref

Bases	Sujets	Serveurs	Groupements
AGRICOLA	Agriculture	Dialog, CDP Online DIMDI	Lyon, Montpellier (CD Rom au Cirad)
AGRIS	Agriculture, Agro-alimentaire	DIMDI, Dialog	Lyon, Montpellier (CD Rom au Cirad)
AQUALINE	Eau	Questel-Orbit, ESA/IRS	Aix, Antony, Lyon, Bordeaux, Rennes, Clermont-Montoldre
ASFA	Eau, océanographie, biologie, écologie, pêche	Dialog, DIMDI, ESA/IRS, STN, Ifremer, CAN.OLE	Aix, Antony, Lyon, Bordeaux, Rennes, Clermont-Montoldre, Grenoble, Montpellier (CD-Rom au Cirad)
BIOSIS	Agriculture, biologie, bactériologie, zoologie, écologie	Dialog, DIMDI, ESA/IRS, STN, CDP Online, CAN.OLE,Data-Star Mead Data	Aix, Antony, Bordeaux, Rennes, Grenoble, Lyon, Clermont-Montoldre
CAB ABSTRACTS	Agriculture, Forêt	ESA/IRS, Dialog, DIMDI, CAN.OLE, STN, Data-Star	Aix, Antony, Lyon, Clermont-Montoldre Clermont-Riom (CD- Rom à l'Inra),Rennes, Grenoble, Nogent (ESA au LEGTA des Barres), Montpellier (CD-Rom au Cirad)
CCN	Périodiques	Sunist	tous groupements (Minitel)
CHEMICAL ABSTRACTS	Chimie	STN, CDP Online, ESA/IRS, Dialog, Questel-Orbit, Data-Star, CAN.OLE	Aix, Antony, Lyon, Clermont-Montoldre, Rennes, Bordeaux, Grenoble
CURRENT CONTENTS	Sommaires de périodiques	CDP Online, Dialog, Data-Star	Tous groupements (disquettes et/ou serveur)

EAU DOC	Eau, environnement	ESA/IRS	Aix, Antony, Lyon, Clermont-Montoldre, Bordeaux, Grenoble, Rennes
ECOTHEK	Environnement, eau, agriculture, pêche, écologie, urbanisme		Tous groupements (Minitel)
ELECTRE	Livres		Tous groupements (Minitel)
ENVIROLINE	Environnement, eau, santé	ESA/IRS, Dialog, Questel-Orbit, DIMDI, Data-Star, CAN.OLE	Aix, Antony, Lyon, Clermont-Montoldre, Bordeaux, Grenoble, Rennes
FLUIDEX	Mécanique, ingénierie	Dialog, ESA/IRS	Aix, Antony, Lyon, Clermont-Montoldre, Bordeaux, Grenoble, Rennes
FRANCIS	Sciences humaines, sciences sociales, économie	Questel-Orbit	Aix, Antony, Bordeaux, Grenoble
FRIPES	Politique, économie	Internet, 3617 FRIPES	Tous groupements
JOEL	Droit, réglementation, administration, législation	3615 JOEL5, 3616 JOEL	Tous groupements (Minitel)
PASCAL	Sciences exactes, biologie, agriculture, médecine	Questel-Orbit, ESA/IRS, Dialog, Data-Star	Aix, Antony, Lyon, Clermont-Montoldre, Clermont-Riom (CD-Rom en BU), Rennes, Grenoble, Nogent (ESA au LEGTA des Barres), Montpellier (CD-Rom au Cirad), Bordeaux
REFLEX	Droit, réglementation législation	36 28 00 02 REFLEX 36 29 00 01 JURIPRO-LEX	Tous groupements (Minitel)
RESAGRI	Agriculture, agro-alimentaire	Sunist	Tous groupements (Minitel)
SCAD	Droit, législation, économie, CEE	Eurobases	Grenoble (à l'IEP)
SIGLE	Sciences exactes, Sciences humaines, sciences sociales, CEE	Blaise, STN, Sunist	Tous groupements (Minitel)
TÉLÉTHÈSES	Thèses	36 15 SUNK*THE 36 16 SUNIST*THE	Tous groupements (Minitel)
VELLEDA	Forêt	36 16 VELLEDA	Tous groupements (Minitel)

Rappelons enfin qu'Internet permet d'accéder à d'autres bases et serveurs. L'utilisation de moteurs de recherche permet de trouver ces adresses.

3.4.2.3. Les Current Contents

Il s'agit d'une revue de sommaires avec résumés d'articles reçue périodiquement sur disquettes (ou sur serveur).

L'organisme éditeur, l'ISI, passe ainsi en revue chaque année 7000 revues et 2000 ouvrages. Chaque fiche comprend une description signalétique complète de l'article cité. L'ISI a regroupé les thématiques des différentes revues avec la dénomination de "séries".

Pour 1996, le Cemagref est abonné dans son ensemble à deux de ces séries :

- Agriculture, biology & environmental sciences
- Engineering, computing & technology

Ces séries sont installées sur les serveurs des groupements et sont donc consultables de n'importe quel PC connecté au réseau. Il suffit au préalable d'y installer l'icône adéquate, ce qui est rapide (s'adresser aux documentalistes ou au responsable informatique).

- Une précision importante sur les Current Contents : il ne s'agit pas de bases de données au sens classique, mais d'un outil de suivi d'actualité, conçu pour être consulté régulièrement. La mise en place de profils de recherche permet de filtrer l'information, et donc de gagner du temps.

- Les astuces appréciées :

- On peut éditer des demandes de tirés à part, les imprimer directement sur des formulaires du type carte postale, lesquels contiennent automatiquement l'adresse de l'auteur et du demandeur. Il n'y a plus qu'à timbrer.

- Les Current Contents permettent d'enregistrer des références bibliographiques sous divers formats informatiques. Donc, à partir des Current Contents, un chercheur peut se constituer une petite base bibliographique personnelle sans ressaisir les notices.

Ici encore, les documentalistes sont à la disposition pour tout conseil sur cet outil, notamment pour établir un profil de recherche.

3.4.2.4. Les serveurs Web

Les serveurs Web sont accessibles car le Cemagref adhère à Renater.

La principale difficulté quant à leur utilisation réside en fait dans l'abondance de l'offre. Le serveur Web du Cemagref comporte un petit cours sur les moteurs de recherche, ainsi qu'un accès aux principaux moteurs.

3.4.2.5. Le serveur Web du Cemagref

Il est accessible à l'adresse suivante : <http://www.cemagref.fr>

Il se présente en deux pages principales:

- la première est accessible à tous :
 - elle offre une présentation de l'Établissement : organigramme, implantations régionales, les activités et Départements, les colloques, ...
 - elle met à la disposition des demandeurs externes deux bases de données : l'une sur les publications des chercheurs (Cemagref publications cité ci-dessus), l'autre sur les produits édités et vendus par le Cemagref.

- la seconde est réservée au domaine Cemagref. On peut notamment y consulter :

- l'annuaire des personnels du Cemagref,
- l'ensemble des bases documentaires internes des groupements,
- les titres de périodiques disponibles au Cemagref.

3.5. Savoir interroger les banques de données

L'interrogation des bases de données en ligne requiert un certain professionnalisme qui est généralement plutôt du domaine de compétence des documentalistes; néanmoins les pages qui suivent tentent d'en dégager les grandes lignes afin d'offrir à tous une certaine autonomie.

Les lignes ci-après reprennent les grands principes de recherche sur bases de données; on pourra également se référer à une étude de cas pratique dans la partie 4 de ce document, à la fin de laquelle on a résumé la stratégie.

3.5.1. Les étapes de recherche

Les grandes étapes de toute recherche documentaire en ligne peuvent se schématiser comme suit :

1. Identification du sujet
2. Formulation du sujet
3. Construction de la stratégie de recherche
4. Exécution de la recherche
5. Évaluation de l'information trouvée
6. Compilation de l'information
7. Diffusion de l'information

3.5.1.1. Identifier le sujet

3.5.1.1.1. Deux techniques préalables

Deux techniques sont utilisables : l'entrevue et le brainstorming.

3.5.1.1.2. L'entrevue :

- elle vise à recueillir les informations techniques,
- elle permet de bien comprendre le sujet et la question, ainsi que leur domaine thématique,
- de bien appréhender le vocabulaire qui l'entoure,
- elle permet de déterminer les limites : temporelles, géographiques, linguistiques,...

Ce dialogue est aussi l'occasion d'expliquer au chercheur les principes de la recherche; on peut reformuler les concepts de plusieurs façons pour s'assurer que la question soit comprise par le moteur de recherche, avec l'accord du chercheur sur les termes. En bref, il est fructueux d'associer le chercheur à l'élaboration de la stratégie de recherche.

3.5.1.1.3. Le brainstorming

C'est la technique qui consiste à réunir plusieurs personnes et à noter tout ce qu'elles expriment sur un sujet donné. En recherche documentaire, elle a l'avantage de faire sortir des concepts sous des formulations diverses; on peut ensuite en faire un réseau sémantique, hiérarchiser les termes, et obtenir ainsi, par exemple, une liste de mots-clés avec leur synonymes.

3.5.1.1.4. Comment identifier le sujet?

a. Définir le contexte de la recherche

À quelle situation se rapporte-t-elle? Pourquoi dois-je la faire?

Ex : situer l'industrie du livre en Europe

b. Identifier l'objet de la recherche

Sur quoi porte l'information recherchée? Qu'est-ce qui m'intéresse? Quelles sont mes connaissances sur le sujet?

Ex : je cherche des informations sur le nombre de livres commercialisés en Europe

c. Délimiter la recherche

Quelles en sont ses limites, notamment géographiques, linguistiques ou autres

Ex : nombre de livres commercialisés ces 3 dernières années dans les pays de la C.E., éventuellement les anciens pays de l'Est, mais à l'exception de la France

d. Préciser le genre de réponses désirées

Quel est le **format** de réponse ou le **type** d'information souhaité

Ex : je cherche une information statistique sous forme de références ou de texte intégral

3.5.2. Formulation du sujet

Une fois le sujet identifié, la formulation consistera à établir un ensemble de mots-clés pertinents qui seront soumis au moteur de recherche des bases de données dans le but d'obtenir une réponse adéquate.

a. Dégager les concepts du contexte

Quels mots ou expressions traduisent le mieux le concept de la recherche?

b. Traduire les concepts en mots-clés

Quels sont les mots-clés qui traduisent le plus adéquatement les concepts?

Ex : Livre / Industrie / Vente / Statistique

c. Déterminer les équivalences

Quels sont les synonymes, équivalents ou voisins?

Ex : Livre ou Livres / Vente / Industrie ou Edition / Statistiques

Traduire ces termes en anglais : Book / Market / Industry / Editing / Statistics

d. Regrouper les mots-clés et les équivalences

(Livre ou Book)

(Vente ou Market)

(Statistique ou Statistics)

Etc...

3.5.2.1. Construction de la stratégie de recherche : les opérateurs de recherche

3.5.2.2. Les opérateurs booléens

La logique booléenne résulte de l'application de l'algèbre de Boole. Elle permet d'opérer trois types de liaisons entre les descripteurs, afin de développer ou au contraire limiter la recherche :

- relation d'intersection qui utilise l'opérateur **ET**

Elle permet de préciser la recherche en combinant deux termes; on retrouvera tous les documents contenant à la fois le terme 1 et le terme 2.

- relation d'union qui utilise l'opérateur **OU**

Elle permet de relier deux descripteurs dont l'un ou l'autre, ou les deux, doivent se trouver dans l'indexation du document pour que celui-ci "sorte" à l'interrogation.

- relation d'exclusion qui utilise l'opérateur **SAUF**

Elle permet de relier deux descripteurs dont le premier doit être présent et le second absent dans l'indexation du document pour que celui-ci "sorte"; elle permet d'éliminer un concept non pertinent

- les parenthèses : elles sont le complément logique des opérateurs booléens. Elles servent à enchaîner plusieurs requêtes dans la même question en évitant les risques de conflit.

(Cf annexe 1)

3.5.2.3. D'autres opérateurs ou procédures de recherche

- la pondération. Elle consiste à affecter à l'indexation comme à la recherche un coefficient de pondération à chaque descripteur, calculé selon son importance dans le document et dans la question

Ex : Livre(3) / Industrie(3) / Vente(2) / Statistique(3)

Le pourcentage de pondération n'a aucun effet sur le choix des documents, mais sur l'ordre dans lequel ils sont triés.

- le voisinage. Il consiste à définir la distance dans le texte des descripteurs sur lesquels porte la recherche; elle s'applique surtout au texte libre. L'opérateur peut demander que deux descripteurs se suivent (1), se trouvent dans une même phrase (2), ou ne soient pas séparés l'un de l'autre par plus d'un nombre déterminé de mots (3).

Ex : (1) la *logique booléenne*

(2) l'algèbre de *Boole* appliqué aux *logiques* de recherche

(3) la *logique* de *Boole*

- la troncature. Elle permet de rechercher un mot par un groupe de lettres, sans tenir compte des suffixes ou préfixes. On distingue la troncature droite ou gauche selon qu'on ignore le suffixe ou le préfixe.

- le masque (ou joker). Il permet de remplacer de 0 à 1 caractère par un symbole (souvent le ?). On recourt à cette section lorsqu'on hésite sur la nature d'un caractère ou si la différence n'est que d'un caractère.

Ex : Francopho?e permettra d'obtenir à la fois les descripteurs "francophone" et "francophobe".

- la comparaison numérique. Elle permet de mener des recherches selon des critères quantitatifs. On parle aussi d'opérateurs de champs numériques.

Ex : on demande des documents publiés à partir de 1990, soit ≥ 1990 .

3.5.3. L'exécution de la recherche

Pendant l'exécution de la recherche, l'expérience du spécialiste en information va très vite permettre à ce dernier de réaliser si ses premières questions sont pertinentes ou pas. Si elles le sont, comme dans le diagramme, on poursuivra la recherche; si elles ne le sont pas, on pourra soit "redresser le cap" en cours de route à l'aide des autres critères élaborés à l'avance, soit se déconnecter et revenir à une meilleure formulation du sujet et construction de la stratégie.

3.5.4. L'évaluation de l'information trouvée

Les ennemis bien connus de toute bonne recherche documentaire sont le silence et le bruit. Ils sont néanmoins utiles en ce qu'ils constituent de bons critères d'évaluation d'interrogation.

3.5.4.1. Le silence

C'est lorsqu'on ne trouve pas de documents, ou alors des documents dont beaucoup ne sont pas pertinents. Un moyen de combattre le silence est d'élaborer une nouvelle requête mieux formulée, comportant des termes équivalents ou plus génériques.

Une stratégie payante consiste à trouver une notice hautement pertinente, en noter les descripteurs et les réutiliser dans une nouvelle requête.

3.5.4.2. Le bruit

On parle de bruit lorsque le résultat est trop important par rapport à la demande. Ce grand nombre représente une perte de temps pour le demandeur. Diverses mesures sont efficaces pour réduire le bruit et ce nombre de documents non pertinents :

- ajouter une limite chronologique,
- employer un terme moins générique et plus spécifique,
- on peut combiner les résultats complets déjà obtenus à un autre terme pour préciser le sujet (à l'aide de ET).

Donc après cette présentation un peu générale de la recherche documentaire, on va aborder ce qui est plus spécifique aux bases de données, et spécialement les bases de données commerciales en ligne.

4. BASES DE DONNÉES : LES PRINCIPAUX MÉTHODES ET OUTILS

4.1. Rappel de la problématique

Tout d'abord il convient de relativiser la notion de faible exploitation évoquée dans l'introduction générale; un tableau figurant dans le Guide de la veille montre une consommation non négligeable de cette source d'information.

Néanmoins l'enquête sur la fonction documentaire de 1995 faisait ressortir que les chercheurs mesurent mal ce que l'on peut tirer des bases de données externes, et qu'ils souhaitent une information élargie. Il s'agit donc plutôt d'une certaine méconnaissance de cet outil de la part des chercheurs et sans doute parfois des documentalistes du Cemagref.

On s'est demandé s'il y avait des moyens de remédier à cette sous-exploitation à travers plusieurs grandes questions:

- comment mieux connaître ces bases commerciales, c'est-à-dire leur existence éventuelle et leur contenu?
- une fois connues, comment juger de leur pertinence?
- peut-on dégager une méthode générale qui puisse s'appliquer à la plupart des interrogations, à la fois sur l'analyse de la pertinence et sur l'organisation du travail?

Une telle méthode, ainsi que les critères d'évaluation de pertinence, pourraient être diffusés auprès de tous les documentalistes du *Cemagref*, et enrichis par la mise en commun.

L'objet principal de cette partie est donc l'utilisation des « bases de données commerciales en ligne ». L'expression est entre guillemets car on a volontairement exclu ici les autres sources d'information, notamment celles accessibles par Internet. On a choisi de se consacrer aux bases de données accessibles par abonnement payant aux grands serveurs internationaux (essentiellement Dialog pour ce document). Par la suite, on emploiera l'expression « bases de données » par simplicité, mais c'est bien des bases de données commerciales en ligne qu'il s'agit.

Cette partie porte, dans un premier temps, sur l'examen des principaux outils et moyens « d'extraire » l'information de ces bases.

Elle est ensuite consacrée à une étude de cas, c'est-à-dire une interrogation réelle répondant à un besoin, où on décrit la méthode qui a permis d'obtenir des résultats.

Enfin dans la section intitulée synthèse et perspectives, on tente de dégager une méthode opérationnelle plus générale sur le plan de l'organisation du travail.

4.2. Présentation des principaux outils de recherche d'information

La problématique d'une recherche documentaire sur base de données ne diffère pas d'une recherche documentaire classique. En général, on a :

- d'une part une collection de bases de données accessibles via un serveur;

- d'autre part un chercheur avec un besoin, une recherche d'information.

Les serveurs qui proposent ces « collections de bases de données » proposent également des outils d'aide sur papier et en ligne.

4.2.1. Les outils sur papier

- catalogue descriptif des bases proposées;
- descriptif complet de chaque base;
- listes de revues prises en compte;
- thesaurus pour certains serveurs ou bases;
- répertoire de base de données du type Gayle's Databases.

4.2.2. Les outils en ligne

- base de données de bases
- données en ligne sur les bases et leur contenu
- outils d'interrogation multi bases.

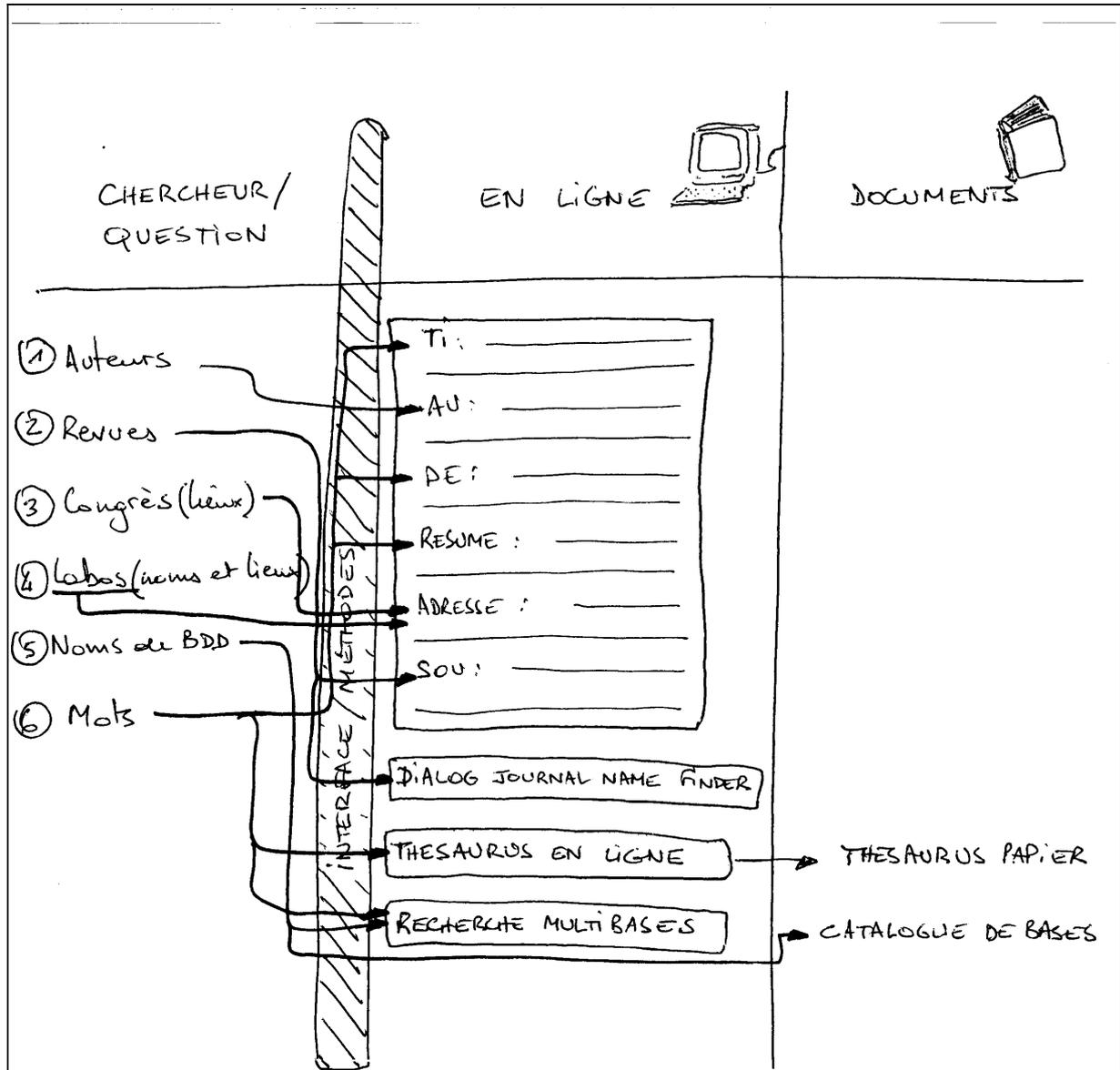
4.2.3. Qu'est-ce qui caractérise la recherche d'information du chercheur ?

On peut appréhender ce que sait déjà le chercheur (ses sources connues d'information) en quelques points-clé :

- des auteurs
- des mots (= des mots-clé)
- des revues
- des lieux de congrès
- des lieux d'organismes ou laboratoires
- des noms d'organismes ou laboratoires

4.2.4. Diagramme synthétique

Cette problématique peut se résumer dans le diagramme suivant :



Les liens dessinés ci-dessus, qui vont des informations connues des chercheurs à celles contenues dans les bases de données, sont décrits dans les pages qui suivent.

4.3. Développement des méthodes et outils

Les exemples qui illustrent les pages ci-après proviennent tous du serveur Dialog, utilisé en mode Commandes. Nous verrons quelles sont les commandes les plus intéressantes, pour chacun des points 1 à 6 du diagramme.

4.3.1. Les auteurs

Les auteurs connus des chercheurs sont donc référencés dans le champ auteurs des bases de données.

- La commande RANK, qui sera décrite plus en détail ci-dessous, permet de montrer qui publie le plus;
- La recherche par auteurs permet d'obtenir des informations sur des noms de revues et de labos.

4.3.1.1. Commande de base de Dialog: SELECT

Dans chaque base de données, les termes sont organisés dans deux index différents :

- l'index de base, qui comprend tous les champs susceptibles de contenir des descripteurs, comme les champs mots-clé, titre, et tous les champs texte en général;
- les index additionnels, qui contiennent tout le reste, et notamment les champs auteurs, revues, date, organisme éditeur, données chiffrées, etc.

Lorsque la recherche porte sur ce type de champ, on doit faire précéder le terme de recherche par l'intitulé exact du champ :

```
?SELECT AU=ASIMOV, ISAAC  
S1 780 AU=ASIMOV, ISAAC
```

4.3.1.2. Les descriptifs techniques des bases

Pour trouver l'intitulé de ces champs, il faut consulter les feuilles bleues (bluesheets), c'est-à-dire les descriptifs techniques des bases.

On peut également l'obtenir en ligne en utilisant la commande HELP :

La commande `HELP FIELD 144` affichera la liste des champs de la base 144 (Pascal), avec leur intitulé et leur type d'indexation.

Une seconde bonne raison de consulter les descriptifs techniques pour les auteurs est de vérifier comment les noms d'auteurs sont entrés dans la base (prénoms complets ou pas, etc...).

4.3.1.3. Les auteurs et la commande EXPAND

On peut utiliser avec profit pour les auteurs la commande EXPAND. Cette commande développe l'index autour du mot choisi : on peut ainsi détecter les différences de saisies pour un même auteur, ou d'éventuelles fautes d'orthographe. L'exemple suivant montre une recherche sur l'auteur Aaron J. Miller:

```
?EXPAND AU=MILLER, A
Ref  Items  Index-term
E1   1      AU=MILLER-WOOD, D. J.
E2   1      AU=MILLER,
E3   0      *AU=MILLER, A
E4   1      AU=MILLER, A.
E5   2      AU=MILLER, A. CAROLYN
E6   1      AU=MILLER, A. G.
E7   3      AU=MILLER, A. H.
E8   1      AU=MILLER, A. J.
E9   1      AU=MILLER, A. PATRICIA
E10  1      AU=MILLER, A. R.
E11  1      AU=MILLER, A. THERESE
E12  1      AU-MILLER, A.J.
```

Il y a des références pertinentes aux lignes E8 et E12 ; il peut y en avoir d'autres plus loin, on active la commande ?PAGE

```
?PAGE
Ref  Item  Index-term
E13  17    AU=MILLER, AARON J.
E14  1     AU=MILLER, AARON J., ED
E15  3     AU=MILLER, AARON J., ED.
E16  2     AU=MILLER, ABRAHAM
E17  1     AU=MILLER, ABRAHAM H.
E18  9     AU=MILLER, ADAM
E19  1     AU=MILLER, ADAM DAVID
E20  1     AU=MILLER, ADAM W.
E21  3     AU=MILLER, ADAM W., JR.
E22  6     AU=MILLER, ALAN
E23  1     AU=MILLER, ALAN D.
E24  1     AU=MILLER, ALAN N.
```

On va donc pouvoir sélectionner les bons « Miller » en tapant les numéros de ligne adéquats : E8, E12, E13, E14, et E15 :

```
?SELECT E8 OR E12:E15
1      AU=MILLER, A. J.
22     AU=MILLER, A.J.:AU=MILLER, AARON J., ED.
S1     23 AU="MILLER, A. J."OR AU="MILLER, A.J.":AU="MILLER, AARON J., ED."

?TYPE S1/3/ALL
1/3/1 DIALOG(R)File 1:ERIC (c) format only

EJ417693  CE522022
An Assessment of Vocational Education Foundations Curriculum Content and
the Entry-Level Skills Needed by Beginning Trade and Industrial Education
Supervisors.
Miller, Aaron J.
Journal of Industrial Teacher Education;...
```

4.3.1.4.

4.3.1.5. La commande RANK appliquée aux auteurs

Lorsqu'on a terminé une recherche par descripteurs, il peut être intéressant de voir quels sont les auteurs qui ont le plus publié sur ce thème. La commande RANK est donc une commande de tri, que l'on peut appliquer au champ auteurs; elle affichera le contenu de ce champ par ordre décroissant d'occurrences.

4.3.2. Les revues

- Les chercheurs connaissent très souvent des titres de revues; le lien du diagramme pointe vers un champ de type source.

4.3.2.1. Les outils en ligne

4.3.2.1.1. Les bases de données de revues

Ces bases de données de revues conviennent pour les recherches par auteurs ou mots-clé; citons notamment Ulrich's International Periodicals Directory.

4.3.2.1.2. Le Dialog Journal Name Finder

Dialog offre une possibilité : le Dialog Journal Name Finder, qui permet de trouver quelle base contient le nom de la revue recherchée. L'exemple qui suit montre la recherche d'un article spécifique de la revue « Natural Health » sur l'utilisation de l'acupuncture pour arrêter de fumer. On doit donc d'abord repérer quelle base indexe cette revue.

```

File 414:Dialog Journal Name Finder 1994/Jun
(c) 1994 Dialog Information Svcs.
SELECT the journal name using the (W) operator and TYPE the record(s):
?SELECT NATURAL(W)HEALTH
      1900 NATURAL
      12892 HEALTH
S1    3 NATURAL(W)HEALTH
...
1/9/1 DIALOG(R)File 414:DIALOG Journal Name Finder(TM) (c) 1994 Dialog
00458780
NATURAL HEALTH (JN=)
  DIALOG FILE 149: HEALTH PERIODICALS DB(TM) 1976-1994/MAY W5
This file contains FULLTEXT records.
  Number of Records for this Journal, 07 JUNE 1994: 227

1/9/2 DIALOG(R)File 414:DIALOG Journal Name Finder(TM) (c) 1994 Dialog
Info.Svcs, All rts. reserv.

00457677
EAST WEST NATURAL HEALTH (JN=)
  DIALOG FILE 149: HEALTH PERIODICALS DB(TM) 1976-1994/MAY W5
This file contains FULLTEXT records.
  Number of Records for this Journal, 07 JUNE 1994: 195

1/9/3 DIALOG(R)File 414:DIALOG Journal Name Finder(TM) (c) 1994 Dialog
Info.Svcs, All rts. reserv.

00290972
NATURAL HEALTH (JN=)
  DIALOG FILE 211: NEWSEARCH(TM) 1994/JUN 06
This file contains BIBLIOGRAPHIC records.
  Number of Records for this Journal, 07 JUNE 1994: 24

```

Les références ci-dessus montrent que les bases 149 et 211 indexent cette revue, que le texte intégral est disponible dans la base 149, et les références bibliographiques dans la base 211. On peut donc lancer ces bases (commande BEGIN) et y lancer la stratégie de recherche :

```
SELECTSMOK? AND ACUPUNCTURE AND JN=NATURAL HEALTH.
```

4.3.2.2. Champ revues et Commande SELECT

Comme pour les auteurs, il faut connaître l'intitulé du champ :

```
? SELECT JN=NEWSWEEK
      S2   811   JN=NEWSWEEK
```

4.3.2.3. Champ revues et Commande EXPAND

Là encore, comme pour les auteurs, cette commande peut s'avérer précieuse pour repérer des titres de revues à partir d'une revue connue.

```
?EXPAND JN=PC WEEK

Ref   Items  Index-term
E1    4       JN=PC PRODUCTS
E2   151    JN=PC TECH JOURNAL
E3  2955    *JN=PC WEEK
E4   145    JN=PC WORLD
E5   87     JN=PERSONAL COMPUTING
E6   59     JN=PUBLISH!
E7  105    JN=RAINBOW
E8   12    JN=ROBOTICS TODAY
E9   13    JN=SCIENCE OF COMPUTER PROGRAMMING
E10  57     JN=SEYBOLD OUTLOOK ON PROFESSIONAL COMPUTING
E11  2      JN=SEYBOLD REPORT ON OFFICE SYSTEMS
E12  38     JN=SIMULATION
```

4.3.3. Lieux de congrès

- Les chercheurs connaissent très souvent le lieu d'un congrès. La zone pertinente des bases de données ont des champs de type « adresse »

4.3.3.1. Les outils en ligne

Le lieu est toujours mentionné : interroger le champ adéquat. Pour trouver le champ adéquat, consulter les Bluesheets.

Il existe des bases de données sur les congrès comme MEETING (Questel) ou FAIRBASE (Dialog).

4.3.4. Lieux et noms de labos

4.3.4.1. Les outils en ligne

Lorsque les chercheurs les connaissent, on trouve les lieux et noms de labos dans les champs de type Adresse et Éditeur; pour avoir l'intitulé de ces champs, consulter les Bluesheets ou utiliser la commande Help Field.

Souvent on connaît mal à priori les noms et adresses de labos. Un bon moyen d'en obtenir est de partir d'un lot de références obtenues par d'autres critères (descripteurs, auteurs,...) et de lancer une commande de type RANK sur le champ adéquat. On peut ainsi visualiser quel labo publie le plus sur une thématique donnée.

4.3.5. Les noms de bases de données

4.3.5.1. Les répertoires de bases de données

Le plus connu est le Gayle's Directory, qui existe à la fois sous forme papier et base de données.

- Pour la forme base de données :
 - utiliser les thèmes plutôt que les mots-clés;
 - ne pas utiliser de noms de composés mais plutôt des classes par exemple «composite» plutôt que «Kevlar».
- Les réponses indiquent :
 - les base de données;
 - les serveurs qui les proposent;
 - éventuellement l'existence d'un CD-Rom.

4.3.5.2. La recherche multibases sur serveur

La recherche multibases (base DIALINDEX sur Dialog) est l'une des meilleures méthodes pour trouver des bases de données sur une thématique donnée. La stratégie de base est de :

- lancer la requête et afficher une liste de base de données qui y répondent;
- les trier par ordre décroissant en fonction du nombre de réponses (RANK FILES);
- les sélectionner en fonction des descriptifs techniques.

4.3.6. Les mots-clé

4.3.6.1. La recherche par mot-clé dans les bases de données

Bien que les progrès en informatique soient quotidiens, notamment au niveau des logiciels et moteurs de recherche, une recherche sur bases de données commerciales comporte toujours les mêmes grandes étapes :

- une étape qu'on peut appeler « sémantique », où on développe l'idée pour obtenir des concepts puis des mots-clé
- une étape plus technique où on teste ces mots-clé dans une stratégie qu'on est souvent amené à développer et à affiner en ligne.

4.3.6.1.1. L'étape « sémantique »

4.3.6.1.1.1. Le dialogue

Par exemple, une équipe s'adresse aux documentalistes et souhaite une recherche complète sur « l'anguille ».

Après dialogue, on s'aperçoit que le thème exact de la recherche est « la migration de l'anguille, avec comme objectif à terme de savoir si l'aquaculture est possible en Europe ».

Le dialogue a donc servi à définir un concept mieux développé et comportant des limites beaucoup plus nettes ; il reste à trouver les mots-clé français et anglais, pour les agencer dans une stratégie efficace.

4.3.6.1.1.2. Concepts et mots-clés

Le sujet : une courte phrase qui décrit l'information que l'on recherche. La plupart des sujets impliquent de 2 à 4 concepts différents qui sont combinés durant la recherche. Le sujet « impact du tabac sur les maladies cardio-vasculaires » comporte deux concepts : « tabac » et « maladies cardio-vasculaires »

Le concept : une idée identifiable à l'intérieur d'un sujet. Par exemple, « tabac » est un concept du sujet « impact du tabac sur les maladies cardio-vasculaires ». Plusieurs mots-clé différents peuvent être utilisés pour exprimer un concept, par exemple ici « tabac », « tabagie », « cigarettes », « fumer », etc....

La stratégie de recherche : les étapes qui conduisent une recherche et permettent de localiser l'information recherchée sur un sujet donné.

Le mot-clé : (keyword ou search term) un mot ou terme que l'on « entre » pour obtenir des références (records) reliées au sujet d'intérêt.

« set » ou lot : le groupe de références créé à chaque fois qu'une recherche est lancée à partir d'un mot-clé donné, ou d'une combinaison de ceux-ci. Chacun de ces lots (set) se voit attribuer un numéro (set number).

À ce niveau, c'est-à-dire avant d'être en ligne, il est intéressant de consulter les thésaurus papier et des dictionnaires (vocabulaire français/anglais).

4.3.6.1.1.3. Les autres paramètres

Ils sont aussi déterminants pour élaborer une stratégie de recherche.

- Le coût : il peut être un élément clé. L'étude de cas du chapitre suivant en fournit un exemple parlant.
- Quel est le degré de « fraîcheur » recherché ? c'est-à-dire sur combien d'années va porter la recherche ?
- Sélectionne-t-on les langues des documents ?
- Quelle est la nature de l'information utile ? Définitions, adresses, statistiques, document de synthèse, ...
- Le degré d'urgence : cela peut influencer sur le choix des bases;
- Privilégie-t-on la facilité de commande des documents ?

- Niveau d'approfondissement : a-t-on besoin d'un simple coup d'œil sur la question ou d'une recherche plus exhaustive sur plus d'aspects du sujet et davantage de bases interrogées ?

4.3.6.1.2. L'étape « technique »

Une fois qu'on a défini une liste de termes potentiels, et qu'on s'est bien mis d'accord sur la stratégie de recherche, on est prêt à se connecter.

4.3.6.1.2.1. stratégie de base

- une fois définis le sujet et ses concepts de base, les écrire;
- pour chacun des concepts, penser aux synonymes, équivalents, orthographes différentes, etc...
- commencer la recherche, un concept à la fois. L'exemple « tabac et maladies cardiovasculaires » que l'on a vu plus haut donnerait la requête suivante :

BEGIN 802	(sélectionne la base de données)
SELECT SMOK? OR TOBACCO	(1er concept)
SELECT HEART (W) DISEASE? OR HEART (W) ATTACK?	(second concept)
SELECT S1 AND S2	croise les 2 concepts

4.3.6.2. La recherche par mots-clé: la commande SELECT de Dialog et ses possibilités

4.3.6.2.1. Rappel sur l'index de base

Lorsqu'on soumet des termes comme dans l'exemple ci-dessus, la recherche se fait par défaut dans ce qu'on appelle l'index de base, à savoir tous les champs pouvant contenir des termes descripteurs; dans la plupart des bases, ce sont les champs Titre, Descripteurs, et les champs « texte » comme le résumé.

4.3.6.2.2. Les troncatures et opérateurs logiques

L'exemple ci-dessus montre qu'on a utilisé des opérateurs logiques et des troncatures. On ne reviendra pas sur ET, OU, SAUF qui sont bien connus. Par ailleurs, les serveurs comme Dialog offre un grand choix pour les troncatures et les opérateurs de proximité :

4.3.6.2.2.1. les troncatures

- la troncature ordinaire (le point d'interrogation) permet d'obtenir tous les mots qui ont la même racine

SELECT SMOK?	donne	smoke smoky smoker smokers smoking smokeless etc
--------------	-------	--

- les troncatures à longueur contrôlée indiquent le nombre maximal de caractères pouvant suivre la racine

SELECT SMOK? ?	smoke smoky ?espace?	on obtient la racine + 1 lettre
SELECT SMOK??	smoke smoky smoker	on obtient la racine + 1 à 2 lettres
SELECT SMOK???	smoke smoky smoker smokers smoking	on obtient la racine + 1 à 3 lettres

- les troncatures à l'intérieur des mots (embedded) sont très utiles par exemple pour chercher à la fois une orthographe anglaise et américaine

SELECT IONI?ATION	ionization ionisation
SELECT FIB??BOARD	fiberboard fibreboard

4.3.6.2.2.2. les opérateurs de proximité

Syntaxe	Exemple	Fonction
(W)	S SOLAR(W)ENERGY	Termes immédiatement adjacents l'un de l'autre, et dans l'ordre indiqué
(nW)	S SOLAR (3W)ENERGY	Termes proches (jusqu'à n mots) l'un de l'autre, dans l'ordre spécifié
(N)	S SOLAR(N)ENERGY	Termes adjacents, mais dans n'importe quel ordre
(nN)	S SMOG(2N)AIR(5N)QUAL?	Les termes doivent apparaître dans une "fenêtre" indiquée, dans n'importe quel ordre. Dans l'exemple ci-contre, on demande que les 3 termes soient présents dans une fenêtre de 10 mots (ces 3 termes + les 7 termes intermédiaires = 1 fenêtre de 10 mots)
(L)	S SOLAR(L)ENERGY	Les termes doivent être dans la même unité de descripteurs telles qu'elles sont définies dans la base de données
(S)	S SOLAR(S)HEAT	Les termes doivent être dans la même unité de sous-champ telles qu'elles sont définies dans la base de données

NOT	ECONOMIC(NOT W)RECOVERY	NOT n'est pas un opérateur de proximité, mais il peut se combiner avec ceux-ci.
-----	----------------------------	--

4.3.6.3. Les autres commandes : EXPAND et RANK

- la fonction de type **EXPAND** : elle décrit l'environnement sémantique d'un terme tel qu'il figure dans les index; elle est intéressante lorsqu'on manque de vocabulaire. Elle a été décrite plus haut pour les auteurs, elle fonctionne de la même façon pour les descripteurs. Elle fournit deux termes au-dessus du terme choisi, plus autant de termes que l'on souhaite en dessous.

-la commande statistique (**RANK**) : il s'agit d'une extraction des termes d'un champ et présentation par ordre alphabétique ou de fréquence. Appliquée aux descripteurs, elle sert à identifier le vocabulaire utilisé par la base de données, et éventuellement peut être un indicateur de pertinence par le nombre d'occurrences.

On utilisera plutôt la première lorsqu'on connaît très peu le vocabulaire de la base, pour en apprendre d'avantage. On emploiera la seconde lorsqu'on a déjà posé des questions, pour découvrir par exemple quel sont les descripteurs les plus utilisés pour un lot particulièrement pertinent.

L'étude de cas pratique qui va suivre fournit un exemple de l'utilisation de RANK sur des descripteurs.

4.4. ÉTUDE DE CAS

Nous allons commenter les méthodes et stratégies d'une recherche documentaire effectuée pour une équipe du Cemagref d'Aix-en-Provence.

4.4.1. Contexte de l'étude

Le Cemagref d'Aix-en-Provence a édité un « Guide technique du forestier méditerranéen français », dont un des volumes traite, en 18 fiches, de la protection des forêts contre l'incendie. Devant le succès de ces fiches, dont l'utilisation dépasse actuellement le cadre strictement français pour lequel elles ont été conçues, un organisme international externe a proposé au Cemagref de lui confier leur adaptation à l'ensemble des pays méditerranéens.

4.4.2. L'étude bibliographique

Parmi les travaux à réaliser par l'équipe du Cemagref figure une étude bibliographique.

Cette étude bibliographique est déjà entamée selon deux axes :

- une première bibliographie provient de notre base de données interne, qui comporte un assez grand nombre de références sur ce thème, de provenances géographiques variées;
- une seconde bibliographie internationale nous a été envoyée par l'organisme en question. Elle provient d'une part de leur base interne, et d'autre part de bases internationales auxquelles cet organisme a des accès privilégiés. Le problème de ce second corpus est qu'il représente 300 références « en vrac », non triées, simplement transférées sur un traitement de texte.

Nos chercheurs ont donc souhaité compléter ce travail par une recherche sur les grandes bases internationales, y compris anglo-saxonnes, pour avoir une vue d'ensemble des publications récentes sur le sujet.

4.4.3. La préparation de l'interrogation

Une première réunion a permis de cerner les thèmes recherchés. Le premier problème à apparaître est l'importance en nombre de ces thèmes; ils figurent ci-dessous :

• **Thème 1: Le mécanisme du feu et son comportement.**

- Combustion et énergie dégagée
 - * La combustion.
 - * Le combustible et ses caractéristiques
 - * Le comburant.
 - * L'énergie dégagée - Quelques chiffres.
- Transfert de chaleur et propagation du feu.
 - * Les modes de transfert de chaleur (conduction, convection, rayonnement).
 - * Les modes de propagation du feu (au sol, feux de cimes, sautes de feu).
- Inflammabilité et combustibilité.
 - * Définition de l'inflammabilité - Méthodes de mesures - Quelques données
 - * Définition de la combustibilité - Méthodes de mesures - Quelques données

• **Thème 2 : Les causes des incendies de forêts.**

- Bases de données sur les incendies de forêts.
 - * intérêt des bases de données.
 - * Création d'une base.
 - * Les bases existantes.
- Origines des incendies de forêt.
 - * Causes naturelles.
 - * Causes humaines involontaires.
 - * Causes humaines volontaires.
- Recherche des causes.
 - * Méthodes de recherches des causes.

Thème 3 :L'analyse du risque.

- Estimation dans le temps - Périodes à risques.
 - * Les indices de risque.
 - * Apports de la météorologie.
 - * Utilisation de la télédétection.
- Estimation dans l'espace - Cartographie.
 - * Les indices de risque.
 - * Utilisation des Systèmes d'information Géographique.
- Modèles de propagation.
 - * Principes des modèles de propagation.
 - * Les modèles existants, leurs utilisations et leurs limites.

• **Thème 4: Prise en compte du risque - Les moyens de la prévention**

- Les principes de la prévention
 - * Éviter le départ de feux.
 - * Limiter la progression du feu.
- Dispositif de surveillance.
 - * Moyens de détection.
 - * Moyens de transmission.
- Cadre législatif.
 - * Lois et décrets.
 - * Mise en application.
- Sensibilisation du public.
 - * Les médias.
 - * La présence sur le terrain.
 - * L'édition de documents.
 - * La formation.

Thème 5: Prise en compte du risque - L'aménagement du territoire.

- Détermination des objectifs de protection.
 - * Hiérarchisation des vulnérabilités.
- Les équipements de protection et leur entretien.
 - * Routes, pistes et zones d'appui.
 - * L'approvisionnement en eau.
- Méthodes de débroussaillage et d'entretien.
 - * Pourquoi débroussailler?
 - * Les différents types de débroussaillage (autour des constructions, en "calissons"...

- * Les différentes techniques de débroussaillage (débroussaillage mécanique, entretien chimique, entretien pastoral, le brûlage dirigé).
 - La structuration de l'espace - Les coupures de combustible.
 - * la compartimentation de l'espace.
 - * Les coupures de combustible.
 - Sylviculture et protection contre l'incendie.
 - * Effet du couvert forestier sur les risques d'incendie.
 - * Sylvicultures.
 - Plan d'aménagement du territoire.
 - * Intérêt d'un plan d'aménagement.
 - * Les différentes formes possibles d'un plan d'aménagement
- **Thème 6: Les principes de la lutte**
- Moyens et techniques de lutte
 - * Moyens humains.
 - * Moyens matériels.
 - * Techniques de lutte
 - Organisation des moyens
 - * Les structures.
 - * la gestion des situations de crise
- **Thème 7 : La réhabilitation des terrains incendiés**
- le diagnostic post incendie.
 - * Impact du feu sur les peuplements,
 - * Impact du feu sur l'environnement (risques d'érosion, pertes économiques, atteintes au paysage...)
 - Les mesures de réhabilitation.,
 - * Les mesures à court terme
 - * Les mesures à moyen terme

Nous avons réduit ces thèmes à une vingtaine de concepts, avec un premier critère de choix qui était le coût; en effet, on souhaitait, au moins dans un premier temps, rester dans une fourchette de 2 à 3000F de coût d'interrogation.

4.4.4. Première étape : l'identification de bases pertinentes

Bien qu'ayant des idées a priori sur des bases de données pertinentes, nous avons préféré une approche systématique, et lancer une recherche multi-bases sur la thématique « incendies de forêts ».

Nous avons donc utilisé la commande DIALINDEX, suivi d'une commande RANK pour lister les réponses par ordre décroissant d'occurrences.

?begin 411

File 411:DIALINDEX(R)

DIALINDEX(R)

(c) 1997, Knight-Ridder Info

Nous n'avons pas recopié ici l'intégralité du traitement. La dernière phrase ci-dessus indique que Dialog a trouvé 319 réponses (donc 319 bases) sur ses 486. Grâce à la commande RANK, et après avoir consulté les descriptifs techniques des bases, nous en avons sélectionné 5 pour la suite du traitement.

Your SELECT statement is:

S FOREST(N) FIRE

Items	File
18	1: ERIC_1966-1997/May
210	2: INSPEC_1969-1997/Jun W4
464	5: BIOSIS PREVIEWES (R)_1969-1997/Jun W4
224	6: NTIS_64-1997/Aug W1
22	7: Social SciSearch (R)_1972-1997/Jun W4
153	8: Ei Compendex (R)_1970-1997/Jul W3
9	9: Business & Industry (R) Jul_1994-1997/Jun 24
990	10: AGRICOLA_70-1997/Jun
7	11: PsycINFO (R)_1967-1997/Jul
3	12: IAC Industry Express (TM)_1995-1997/Jun 23
4	14: Mechanical Engineering Abs_1973-1997/Jul
86	15: ABI/INFORM (R)_1971-1997/Jun W3
144	16: IAC PROMT (R)_1972-1997/Jun 23
17	18: IAC F&S INDEX (R)_1980-1997/Jun 23
217	29: Meteor. & Geoastro. Abs._1970-1997/Mar
50	30: AsiaPacific_1985-1997/May B2
2	32: METADEX (R)_1966-1997/Aug B1
55	35: Dissertation Abstracts Online_1861-1997/Jun
9	37: Sociological Abstr._1963-1997/Jun
22	38: America:History & Life_1963-1997
2	39: Historical Abstracts_1973-1997/Iss48/B1
172	40: Enviroline (R)_1975-1997/May
25	41: Pollution Abs_1970-1997/Jun
1	42: PHARMACEUTICAL NEWS INDEX_1974-1997/Jun W3
14	44: Aquatic Science Abstracts_1979-1997/Jul
64	46: A-V Online_1997/Jun. Q2
438	47: Magazine Database (TM)_1959-1997/Jun 24
9	49: PAIS INT._1976-1997/MAY
811	50: CAB Abstracts_1972-1997/May
Examined 50 files	
2	64: Global Mobility Database (r)_1965-1997/Apr
112	66: GPO Mon. Cat._Jul-1997/Apr
176	68: Env.Bib._1974-1997/Aug
3	69: Energyline (R)_1970-1993/Dec
59	73: EMBASE_1974-1997/Jun W2
5	75: IAC Management Contents (R)_86-1997/Jun W3
106	76: Life Sciences Collection_1982-1997/May
42	77: Conference Papers Index_1973-1997/May
34	78: DIALOG SourceOne (SM) Eng_1991-1996/Oct W4
45	80: IAC Aerospace/Def.Mkts (R)_1986-1997/Jun 24
3	86: Mental Health Abstracts_1969-1997/Jun
360	88: IAC BUSINESS A.R.T.S._1976-1997/Jun 24
84	102: ASI_1973-1997/Apr
175	103: Energy SciTec_1974-1997/May B2
248	108: Aerospace Database_1962-1997/Jun
1	109: Nuclear Sci. Abs._1948-1976
129	111: Natl.Newspaper Index (SM)_1979-1997/Jun 24
5	113: European R&D Database_1997
9	114: Encyclopedia of Associations/Nov 1996_
14	115: Research Centers & Services_1995/Nov
50	117: Water Resour.Abs._1967-1997/May
2	119: Textile Technol.Dig._1978-1997/Jun
56	120: US Copyrights_1978-1997/Jun 16
1	133: S&P's Corp.Descrip.+News_1997/Jun B2
Examined 100 files	
25	135: Congress.Rec.Abs._1981-1993/Nov
18	137: Book Review Index_1969-1997/Q1
10	142: Social Sciences Abstracts_1983-1997/May
71	143: Biol. & Argic. Index_1983-1997/May
417	144: Pascal_1973-1997/May
223	146: Washington Post Online_1983-1997/Jun 23

990	10:	AGRICOLA_70-1997/Jun
811	50:	CAB Abstracts_1972-1997/May
432	203:	AGRIS_1974-1997/Apr
417	144:	Pascal_1973-1997/May
224	6:	NTIS_64-1997/Aug W1
172	40:	Enviroline(R)_1975-1997/May

4.4.5. Premier essai sur une seule base

Avant de nous lancer dans une recherche multibases, nous avons préféré tester nos concepts et leur traduction en descripteurs dans une seule base, Pascal. Voici le résultat de la commande DISPLAY SET qui permet de récapituler l'historique des questions.

Set	Items	Description
S1	378	FIRE(W)BEHAVIOUR OR FIRE(S)FUEL OR FIRE(S)SPREAD
S2	2087	FLAMMABILITY
S3	140509	COMBUST?
S4	54584	FOREST
S5	66418	FUEL
S6	3021	4 AND S5
S7	4140	3 AND S4
S8	279	S4 AND S5
S9	327	S3 AND S4
S10	170	S8 AND S9
S11	50	FIRE?(N)CAUSE
S12	200	FACTEUR ANTHROPIQUE AND FIRE
S13	0	S11 AND S12
S14	457	FIRE(S)DATA
S15	1346	FIRE? AND DATA
S16	27197	FIRE?
S17	23271	DATABASE
S18	63	S16 AND S17
S19	4	S18 AND S4
S20	15	S4 AND FIRE(S)EVIDENCE
S21	748	FIRE(W)HAZARD
S22	44	S21 AND S4
S23	1502	S4 AND S16
S24	23062	REMOTE SENSING
S25	68	S23 AND S24
S26	30472	CLIMATE
S27	97	S26 AND S23
S28	913331	MODEL?
S29	219	S28 AND S23
S30	21190	CARTOGRAPHY
S31	24	S30 AND S23
S32	27029	METEOROLOGY
S33	17	S32 AND S23
S34	160463	PREVENT?

S35	88	S34 AND S23
S36	12	FOREST (S) FIRE (S) PREVENT?
S37	165060	DETECTION
S38	41	S23 AND S37
S39	996	FIREFIGHTING
S40	56	S39 AND S23
S41	3	FUEL (W) BREAK
S42	8657	SYLVICULTURE
S43	99	S42 AND S23
S44	387099	DIAGNOS?
S45	5	S23 AND S44
S46	5	VEGETATION (W) CLEAR?

Comme certains lots ne nous semblaient pas très pertinents (les parties visualisées semblaient comporter un certain taux de bruit), nous avons souhaité affiner notre connaissance du vocabulaire, notamment anglo-saxon.

Pour cela, nous avons lancé un tri (RANK) sur les descripteurs d'un des lots antérieurs qui paraissait, lui, bien situé dans la problématique générale feux de forêts.

?rank de

Started processing RANK
...Ranking 500 of 1227 records
...Ranking 1000 of 1227 records
Completed Ranking 1227 records
DIALOG RANK Results

RANK: S2/1-1227 Field: /DE File(s): 144
(Rank fields found in 1224 records -- 7260 unique terms) Page 1 of 908

RANK No.	Items	Term
1	963	FORET
2	834	INCENDIE
3	757	FIRE
4	590	INCENDIO
5	564	BOSQUE
6	381	FORESTS
7	339	SPERMATOPHYTA
8	337	FOREST

P = next page Pn = Jump to page n
P- = previous page M = More Options Exit = Leave RANK

To view records from RANK, enter VIEW followed by RANK number, format, and item(s) to display, e.g., VIEW 2/9/ALL.

Enter desired option(s) or enter RANK number(s) to save terms.

?p

RANK No.	Items	Term
9	237	AMERIQUE DU NORD
10	236	ETATS UNIS
11	236	NORTH AMERICA
12	235	UNITED STATES
13	228	AMERICA DEL NORTE
14	228	FEU VEGETATION
15	227	CANOPY FIRE

16	216	ESTADOS UNIDOS
17	207	AMERICA
18	207	AMERIQUE
19	202	AGRICULTURA
20	202	AGRICULTURE
21	198	FUEGO VEGETACION
22	196	SYLVICULTURE
23	179	FORESTRY
24	175	SILVICULTURA
25	171	GYMNOSPERMAE
26	168	CONIFERALES
27	154	EUROPE
28	143	ARBRE FORESTIER RESINEUX
29	142	SOFTWOOD FOREST TREE
30	140	EUROPA
31	122	INCENDIE CONTROLE
32	121	PRESCRIBED BURNING
33	112	ANGIOSPERMAE
34	112	ECOLOGIE
35	111	ECOLOGY
36	108	DICOTYLEDONES
37	108	ECOLOGIA
38	103	INCENDIO CONTROLADO
39	101	DAMAGE
40	99	ARBOL FORESTAL RESINOSO
41	99	CANADA
42	97	DEGAT
43	91	NORDAMERIKA
44	88	AUSTRALIA
45	88	AUSTRALIE
46	85	FRANCE
47	83	PROTECTION
48	81	MEDITERRANEAN REGION
49	81	REGION MEDITERRANEENNE
50	78	PROTECCION
51	75	AGRONOMIA
52	75	AGRONOMIE
53	75	AGRONOMY
54	74	FRANCIA
55	73	SOL
56	66	REGION MEDITERRANEA
57	66	VEGETATION
58	65	DANO
59	64	DYNAMIQUE VEGETATION
60	64	REMOTE SENSING
61	64	TELEDETECTION
62	64	VEGETATION DYNAMICS
63	63	FIRES
64	62	LUTTE INCENDIE
65	62	OCEANIA
66	62	OCEANIE
67	61	FIREFIGHTING
68	61	VEGETACION
69	60	ACTION FEU
70	60	FIRE EFFECT
71	57	VEREINIGTE STAATEN
72	56	ARBRE FORESTIER FEUILLU
73	56	FOREST SOIL
74	56	INCENDIE FORET
75	56	SOL FORESTIER
76	55	HARDWOOD FOREST TREE

77	53	ECOSYSTEME
78	52	DETECCION A DISTANCIA
79	51	MODELE MATHEMATIQUE
80	50	ACCION FUEGO
81	50	COMBUSTIBLE
82	50	SUELO FORESTAL
83	49	FACTEUR MILIEU
84	49	FUEL
85	48	ENVIRONMENTAL FACTOR
86	48	FACTEUR ANTHROPIQUE
87	47	CALIFORNIA
88	47	CALIFORNIE
89	47	INVERTEBRATA
90	46	LITIERE
etc...		

À ce niveau-là nous nous sommes déconnectés pour affiner notre stratégie.

4.4.6. L'affinage de la stratégie

Cela nous a permis de vérifier la bonne traduction anglaise de nos descripteurs. Nous avons changé quelques-uns des termes anglais et nous avons conservé tous les termes qui donnaient des réponses satisfaisantes lorsqu'on listait quelques fiches. Cela a donné la liste suivante, prête à être essayée en recherche sur nos 5 bases :

- Concept feux de forêts :
 - soit le lot résultant du croisement des termes « FOREST » et « FIRE? » pour une approche large
 - soit WILDFIRE pour une approche plus précise
- Puis les autres concepts (une quinzaine) destinés à être croisés avec les premiers, à savoir :
 - FUEL(W)BREAK
 - FIRE(W) CONTROL ou FIREFIGHTING
 - FLAMMABILITY
 - SYLVICULTURE
 - BRUSH(W)CONTROL
 - DATABASE
 - HAZARD
 - REMOTE SENSING
 - CLIMATE
 - MODEL
 - CARTOGRAPHY
 - METEOROLOGY
 - PREVENT?
 - DETECTION
 - DAMAGE ou FIRE EFFECT
 - FIRE(S)EVIDENCE

Un autre problème est alors réapparu : celui du coût. En effet, les réponses de Pascal uniquement nous donnaient déjà un nombre important de références, et il était clair que notre budget serait rapidement dépassé sur les 5 bases.

Nous avons profité de notre première interrogation pour obtenir en ligne les fiches techniques (bluesheets) de nos bases, ainsi que leurs tarifs. Le plus coûteux dans les bases de données est le prix de téléchargement par référence : cela devient donc souvent un critère

prioritaire. Après un rapide calcul, nous avons convenu avec nos chercheurs de nous limiter à une vingtaine de réponses par concept pour rester dans notre budget.

4.4.7. Comment restreindre le nombre de références?

Nous avons utilisé les 4 techniques suivantes :

4.4.7.1. Affiner le vocabulaire :

Lorsque le lot « FOREST(N)FIRE? » donnait encore trop de réponses une fois re-croisé, nous avons utilisé le second , « wildfire », qui comprenait moins d'occurrences.

4.4.7.2. La commande REMOVE DUPLICATE (RD):

Particulièrement efficace en multi-bases, elle élimine les doublons. La fiche qui est conservée est celle de la première base indiquée après la commande BEGIN, il peut être intéressant d'en tenir compte lors des recherches multi-bases. Donc lorsqu'on préfère les formats des fiches d'une base à celles d'une autre, la rentrer en premier après BEGIN. On peut également conserver les fiches éliminées dans un fichier temporaire pour les vérifier.

4.4.7.3. la restriction :

La restriction de la commande **SELECT** à un ou deux des **champs** de l'index de base ;
syntaxe : **SELECT S2/DE, TI**

En effet, la recherche par défaut fonctionne, on l'a dit, sur les index de base qui concernent presque tous les champs texte. Il se produit parfois, notamment lorsqu'on a de longues fiches avec beaucoup de texte, que l'opérateur AND perde de sa force de tri et qu'on ait du bruit. C'est là sa limite. On peut éliminer une partie de ce bruit par la stratégie illustrée par l'exemple ci-dessus; on peut même réduire uniquement au champ descripteurs.

- La restriction par date

Elle est non seulement un critère simple et connu, mais elle a l'avantage de fonctionner à l'identique sur toutes les bases de Dialog.

On peut obtenir la liste des facteurs limitants d'une base par la commande **HELP LIMIT** n° de base. Par exemple, voici ceux de Pascal :

```
LIMITING
Sets and terms may be limited by Basic Index suffixes, i.e., /AB, /BT, /DE,
/DF, /ED, /FD, /SD, /OD, /ET, /FT, /ST, /OT, /TI, as well as the
following features:
  SUFFIX      FIELD NAME/EXAMPLE
/None        Publication Year
              S S2/1987:1988
/ABS         Abstract Present -4
              S S6/ABS
/ART         Journal Article -4
              S S5/ART
/ENG         Original Document in English -4
              S X(W)RAY(W)DIFFRACTION/ENG
/FRENCH      Original Document in French -4
              S DIFFRACTION(W)RX/FRENCH
```

4.4.8. L'interrogation multi-bases

```
?B 40,6,10,50,203,144

SYSTEM:OS - DIALOG OneSearch
File 40:Enviroline(R) 1975-1997/Jun
(c) 1997 Congressional Information Service
File 6:NTIS 64-1997/Aug W4
Comp&distr 1997 NTIS, Intl Copyright All Rights Re
File 10:AGRICOLA 70-1997/Jul
(c) format only 1997 Knight-Ridder Info
File 50:CAB Abstracts 1972-1997/Jun
(c) 1997 CAB International
File 203:AGRIS 1974-1997/Jun
Dist by NAL, Intl Copr. All rights reserved
File 144:Pascal 1973-1997/Jun
(c) 1997 INIST/CNRS
```

On a donc d'abord activé les 5 bases :

Puis on a lancé les 2 premiers concepts de base, destinés à être croisés avec les autres.

```
?s FOREST
S1 470101 FOREST
?S FIRE?
S2 101519 FIRE?
?S WILDFIRE
S3 1332 WILDFIRE
?s S1 AND S2
470101 S1
101519 S2
S4 17801 S1 AND S2
```

Nous avons volontairement conservé ces 2 sets (S3 et S4) pour effectuer les croisements avec les autres concepts :

- en priorité avec le plus complet (S4) qui est aussi le plus volumineux
- et lorsque les croisements ont donné des résultats encore trop volumineux pour être réduits comme on l'a décrit ci-dessus, on utilise S4, concept plus spécialisé, qui restreint beaucoup avec l'opérateur AND.

Nous n'allons pas reproduire ci-dessous toute la recherche, mais 2 exemples, ceux du débroussaillage (brush control) et de la prévention.

```

?s BRUSH (W) CONTROL
    12513 BRUSH
    1924625 CONTROL
    S38 1146 BRUSH (W) CONTROL ==> on a 1146 réponses
?s S38 AND S3
    1146 S38
    1332 S3
    S39 4 S38 AND S3 Croisé avec le premier concept
(wildfire), on a trop peu de réponses; on essaie donc le second
?s S38 AND S4
    1146 S38
    17801 S4
    S40 25 S38 AND S4 Ici plus que 25 réponses
?RD S40 On enlève les doublons :
...completed examining records pas de doublons
    S41 25 RD S40 (unique items)
?s S41/1990-1997
    25 S41
    6338629 PY=1990 : PY=1997 Réduction à la période 1990/1997
    S42 13 S41/1990-1997 plus que 13 réponses

S90 317482 PREVENT?
?s S90 AND S4
    317482 S90
    17801 S4 Trop de réponses avec S4
    S91 1752 S90 AND S4 ==> on lui préfère S3
?s S90 AND S3
    317482 S90
    1332 S3
    S92 122 S90 AND S3
?s S92/1991-1997 On réduit aux 6 dernières années
    122 S92
    5436164 PY=1991 : PY=1997
    S93 40 S92/1991-1997
?s s93/TI,DE
    S94 28 S93/TI,DE On réduit aux champs Titre et Descripteurs
?RD S94 On enlève les doublons (il y en avait 4)
...completed examining records
    S95 24 RD S94 (unique items)

```

On a donc reproduit ce processus autant de fois qu'il y avait de concepts; la répétition pourrait sembler fastidieuse, et on aurait très bien pu, par exemple, éliminer les doublons en fin de procédure. Mais d'une part cela s'est avéré assez rapide, et d'autre part cela nous a permis de maîtriser au fur et à mesure le nombre de documents, et donc le coût.

Ensuite, on a pu rassembler toutes les étapes pertinentes en une seule question, vérifier que le nombre de réponses était correct, et éliminer encore d'éventuels doublons :

```
?s S5 OR S18 OR S33 OR S36 OR S42 OR S51 OR S57 OR S61 OR S69 OR S76 OR
S83 OR S95 OR S99 OR S109 OR S111
      16 S5
      13 S18
      28 S33
       9 S36
      13 S42
      26 S51
      22 S57
      16 S61
      24 S69
      19 S76
      19 S83
      24 S95
      20 S99
      16 S109
      16 S111
S112      217 S5 OR S18 OR S33 OR S36 OR S42 OR S51 OR S57 OR S61 OR
          S69 OR S76 OR S83 OR S95 OR S99 OR S109 OR S111
?RD S112
...examined 50 records (50)
...examined 50 records (100)
...examined 50 records (150)
...examined 50 records (200)
...completed examining records
S113      212 RD S112 (unique items)
```

Enfin on a téléchargé les 212 références, en format complet.

```
?TYPE S113/5/ALL TAG
```

Comme on souhaite intégrer plus tard ces références dans le logiciel bibliographique End Note Plus, on a ajouté l'option « tag »; cette option ajoute les intitulés de champs (tags), qui sont nécessaires à l'édition de filtres d'importation. Mais il était donc nécessaire de le prévoir avant de télécharger les références. La notice de End Link fournit la liste des formats de téléchargement recommandés pour les principaux serveurs et base de données, pp.30-31 (cf en annexe 2).

4.4.9. Interrogation complémentaire

Nos chercheurs ont ensuite souhaité une recherche complémentaire, sur les mêmes bases, mais sur une thématique plus méditerranéenne. On ne détaillera pas ici la stratégie en ce qui concerne les mots-clé; ce qui est intéressant à souligner, c'est qu'on a procédé à une interrogation un peu large (environ 700 références), et qu'au lieu de limiter davantage, on a choisi de visualiser ces références dans un format gratuit (titre seul). Après avoir visualisé les 100 premières, on a volontairement choisi de visualiser, et donc de capturer, les 600 suivantes. Grâce à l'impression papier, les chercheurs ont pu se faire une meilleure idée du contenu de chaque base.

L'inconvénient du format titre est qu'il est court. Son gros avantage est d'être « gratuit », c'est-à-dire qu'il n'y a aucune facturation par référence, mais un simple coût de temps de connexion. De plus les titres sont accompagnés d'un numéro de côte qui permet le cas échéant de retourner rapidement dans la base pour télécharger les références complètes les plus intéressantes. Il existe parfois plusieurs formats gratuits par base; là encore ce sont les Bluesheets qui les décrivent.

Pour mémoire, la capture de plus de 700 titres a coûté moins de 200F de temps de connexion, et les utilisateurs semblaient très satisfaits.

4.4.10. Rappel de la stratégie

- préparation de l'interrogation : liste de concepts et de descripteurs potentiels
- identification de bases pertinentes :
 - DIALINDEX + RANK
 - choix de 5 bases
- essai sur une seule base (Pascal)
 - vérification et affinage du vocabulaire avec la commande RANK DE
 - affinage de la stratégie : limitation à une vingtaine de réponses par concept pour rester dans les coûts
- recherche multibase et application de cette stratégie à tous les concepts
- résultat final : une dernière étape avec 212 références
- téléchargement des références avec les intitulés de champs pour import dans EndNotePlus.

Cette stratégie ne prétend pas s'appliquer à toute interrogation de base de données. Comme toujours dans ce domaine, on doit réagir en fonction de la situation, en opérant des choix à l'intérieur d'une palette d'outils fournis par le serveur. S'il ne s'avère pas indispensable, en général, de connaître toutes les syntaxes de commandes sur le bout des doigts, il est préférable de bien connaître au moins l'existence et les potentialités de ces outils.

La partie suivante récapitule néanmoins les grandes stratégies de base, applicables à toute interrogation, ainsi que les principaux problèmes et solutions pour les résoudre.

5. SYNTHÈSE ET PERSPECTIVES

5.1. Les autres serveurs

Toutes les stratégies et méthodes décrites précédemment ne sont pas limitées au serveur Dialog. Les autres serveurs offrent des fonctionnalités comparables, dont la plupart sont synthétisées dans le tableau ci-après.

DOC315/JPB			
	QUESTEL	DIALOG	ORBIT
8	Voir les résultats		
8.1	Voir un champ précis	..VI TI 1-5	T S1/TI/1-5
8.2	Voir une notice complète	..VI (QU 3) MAX 4-4	T S1/5 /4-4
8.3	Voir toutes les notices	..VI MAX (danger)	T S3/5/all
			PRT TI 165 PRT FU 4-4 SS 2 PRT FU SET
9	Sauver la recherche		
9.1	Temporaire	..ST SV	SAVE TEMP ABCD
9.2	permanente	..SV RE EFGH	SAVE EFGH
9.3	Exécution des profils	..EX ABCD	EXS ABCD
			SAVE ABCD STORE EFGH RECALL ABCD
10	Surveillance automatique		
10.1	Création	..SV PF KLMN	SAVE SDI KLMN
			SDI (répondre aux questions)
11	Utilisation de la mémoire		
11.1	Stockage d'une donnée	..MEM ..MEM /PN RG 1	MAP MAP PN TEMP XY098 (nom de la sauvegarde)
11.3	Changement de base	..BA CABS	B 299
11.4	Exécution	*MEM/PN	EXS XY098
11.5	Visualiser la mémoire	..VI MEM	(via éditeur)
11.6	Effacer ou garder	..EF MEM 6-10,15	
			PRT SEL PRT SEL PN 1-x FILE CASM SEL 1-x SHOWSEL 1-x KEEPSEL 1-5,11,14-15
12	Statistiques		
12.1	Liste des pays des brevets	..MEMT	RANK
12.2	Liste des mnémoniques des sociétés	..MEMT 1-x /PN RG 1 LG 2 ..MEMT /PA RG 1	RANK PC RANK CC
12.3	Liste des sociétés	..MEMT /PA RG 2	RANK PA (sauvegarde demandé)
12.4	Exécution	*MEMT /PA	EXS TD005
			GET RANK GET RANK PC GET RANK CC GET RANK PA SET TOSEL SEL 1-x
13	Paramétrage		
13.1	Temporaire	..OP	SET
13.2	Définitive	..UT	
13.3	Nbr de colonnes	..UT NC	SET SCREEN NC NL
13.4	Relief	..UT TER 1 ou 2	SET HILIGHT ON
			DISPLAY Répondre aux questions de DISPLAY
14	Recherche multibases		
14.1	Se créer un groupe de bases	..CL NOM1 BASEE1 BASE2	B 2,6,8,34,108
14.2	Choisir un groupe	..CL NOM1	
14.3	Poser une question	MOT1 ET MOT2	S MOT1 AND MOT2
14.4	Voir les groupes	..CL	
			POWERINDEX CLUSTER NAME NOM1 BASE1 BASE2 *INDEX* NOMCLUSTER MOT1 AND MOT2 SHOWCLUSTER
15	Commandes diverses		
15.1	Commandes documents		
15.2	Changer les commandes		
15.3	Impressions différés		

5.2. Synthèse des stratégies de base

Ce tableau illustre les méthodes les plus couramment utilisées dans l'interrogation des bases de données, à un niveau général.

D'autres tableaux viendront ensuite compléter certains points de méthode particuliers.

1. Avant la connexion

1.1. Dialoguer ==> amasser un maximum d'informations

règle de base : faire l'interrogation avec le chercheur

But : - description du concept fondamental en une phrase claire

- liste de tous les facteurs limitants :

géographique

langue

Coût

Délais

Fraîcheur de l'info

Nature de l'info

etc

les trier par ordre décroissant d'intérêt avec le demandeur

1.2. Déterminer les bases à utiliser

1.2. Rédiger liste termes pouvant être descripteurs

Français

Et surtout en anglais

Orthographes possibles

synonymes ou contraires

2. Une fois connecté

2.1. Technique de base

Si le sujet est inconnu : interrogation multibases (Dialindex)

Soumettre les termes un par un

Afficher les historiques (Display Set) ==> aide à vérifier stratégie

Les croiser entre eux plutôt vers la fin

Sortir que les titres (gratuit), se déconnecter temporairement et trier sur papier (lorsque le nombre de références le permet)

2.2. Appliquer les facteurs limitants

- en fonction des priorités du demandeur

- en fonction des problèmes rencontrés

L'ordre dans lequel on opère n'a qu'un intérêt relatif dans la mesure on peut toujours revenir sur des étapes antérieures grâce à l'historique

3. Liste des problèmes et de leur solutions

• **auteur pas trouvé** : EXPAND auteur ==> orthographes différentes

• **Journal pas trouvé**: idem

• **Vocabulaire incertain** : EXPAND ==> environnement immédiat dans index

RANK DE==>environnement tous descripteurs sur même thème

• **Bases incertaines** : index de bases (DIALINDEX)

• TROP de données

- affiner vocabulaire (cf ci-dessus)

- restreindre recherche sur un champ (TI ou DE)

- restriction par date

- restriction par langue

- autres restrictions possibles : cf HELP LIMIT n°base

- éliminer doublons

• Pas assez de données

- vérifier vocabulaire : orthographes différentes
 - utiliser troncatures : l'augmentation du nombre de références est important
 - troncatures internes pour orthographes anglaise et américaine
- examiner avec soin l'indexation des documents satisfaisants pour avoir d'autres idées de descripteurs
- opérateurs de proximité : bien que généralement utilisés pour restreindre la recherche, on peut obtenir l'effet inverse : - au lieu de (W), utiliser (2N), (S) ou même AND
- sommes-nous dans la bonne base ? la meilleure des stratégies ne peut rien dans une base inadéquate : tester une stratégie simplifiée dans DIALINDEX, puis faire un tri (RANK) sur ces bases.

5.2.1. Bilan des critères de pertinence

Précisons d'abord le champ : par critères de pertinence, on entend la recherche de bases pertinentes pour répondre à une demande, puis l'estimation en ligne de la pertinence de leur contenu. Si l'on dresse le bilan des techniques utilisées, on obtient les suivantes :

- **lors de l'étape dialogue :**
 - utiliser l'expérience du chercheur et du documentaliste ;
 - s'adresser aussi aux autres chercheurs ou documentalistes susceptibles de connaître des bases ;
- **En ligne :**
 - poser la question sur DIALINDEX (index de bases) pour voir quelles bases répondent ;
 - lancer une commande RANK pour les trier par ordre d'importance de réponses ;
 - pour les bases que l'on ne connaît pas et qui présentent beaucoup de réponses, utiliser les descriptifs techniques pour vérifier la thématique de la base ;
 - visualiser des lots de références, de préférence en format gratuit de type « titre », pour vérifier plus avant la pertinence.

5.2.1.1. limites et intérêt de la méthode

Les limites de cette méthode de travail sont les suivantes :

- on utilise généralement Dialindex avec une question assez large, donc on obtient beaucoup de réponses (bruit) ;
- le critère de tri que l'on utilise avec Rank est quantitatif : ce sont les bases qui présentent le plus de réponses qui s'affichent en premier. Le critère de quantité ne constitue pas forcément un bon critère de pertinence. Par exemple, des bases peuvent présenter peu de réponses, mais pertinentes ;
- c'est pour cette raison que l'on est amené à effectuer un tri qualitatif à travers la lecture des fiches techniques et d'un certain nombre de références pour chaque base.

L'ensemble constitue une sorte de processus qu'on pourrait qualifier d'itératif, où l'on est amené à reproduire un certain nombre de procédures. Son avantage est d'être simple et d'une logique assez naturelle : on avance en enrichissant ses connaissances au fur et à mesure, et on a toujours loisir de revenir sur un point de stratégie grâce à la capture (enregistrement)

des sessions (cette remarque vaut pour l'ensemble des techniques d'interrogation de bases de données). Et malgré la réserve émise plus haut, une commande comme Rank (tri quantitatif) est puissante et rapide, et s'avère globalement efficace y compris comme outil de mesure de pertinence.

5.2.2. L'interrogation multibases : avantages et inconvénients

L'interrogation multibases est souvent une conséquence directe de l'utilisation de Dialindex. Les avantages sont évidents : une fois identifiées plusieurs bases pertinentes, il est commode de lancer la même recherche sur ces bases. C'est également plus rapide, donc a priori moins coûteux.

La procédure présente néanmoins quelques inconvénients :

- l'élimination des doublons semble perdre en efficacité ; l'étude de cas en comportait encore un certain taux ;

- si l'on utilise des facteurs limitants à l'interrogation, il faut veiller à ce qu'ils soient communs aux bases interrogées. Certains, comme la date, sont communs à toutes les bases de Dialog, et d'autres sont spécifiques à certaines bases ;

- lorsqu'on télécharge les références avec les intitulés de champs (format TAG) afin de les importer dans un logiciel de type EndNote, il s'avère assez complexe de fabriquer un filtre d'importation correspondant à tous les formats des différentes bases. Il peut s'avérer plus rapide d'importer les notices en plusieurs phases successives, avec un filtre par base. On doit alors retrier les notices base par base, ce qui n'est pas très long puisque Dialog exporte les notices dans l'ordre où on a ouvert les bases.

5.2.3. Rappel sur les « Bluesheets »

On a très souvent évoqué ces descriptifs techniques tout au long de ce document. Il faut distinguer le descriptif fourni en standard (simple liste avec la thématique générale) de ces feuilles bleues qui décrivent les bases en détail.

Elles sont indispensable notamment lorsqu'on doit affiner les stratégies.

Voici leurs caractéristiques qui servent le plus souvent :

- description fine du champ thématique de la base (= critère de pertinence!) ;
- détail des intitulés de champ, et donc des index autres que celui de base (ex Auteur, source, etc...) ;
- facteurs limitants ;
- formats disponibles et coûts.

Comme il est coûteux et inutile de les commander dans leur intégralité, la procédure la plus intéressante est de récupérer les bluesheets des bases que l'on est amené à interroger sur la base Dialog Bluesheets, pour pouvoir en disposer à tout moment lors des interrogations. La procédure en ligne est rapide et peu chère.

5.2.4. Projet de fiche d'interrogation de bases

Afin d'améliorer l'étape de dialogue et de définition des concepts, voici un projet de fiche de préparation à l'interrogation des bases de données. Cette fiche pourrait être soumise aux autres documentalistes pour discussion et enrichissement éventuel.

Projet de fiche de préparation à l'interrogation de bases de données

Date :	
Nom du chercheur :	
Pour stagiaires : maître de stage :	
Groupement, Division :	
Département :	
Programme :	
Question en texte libre :	
Domaine scientifique général de la question	
Attentes par rapport à la question. À quoi cette recherche va-t-elle servir?	
Sur combien d'années interroger?	
Langues : (à préférer ou à éliminer)	
Zones géographiques éventuelles : (à préférer ou à éliminer)	
Degré d'urgence	
Nature de l'information recherchée (biblio pure, adresses, statistiques,...)	
Coût (budget à respecter)	
Quels sont les mots qui vous paraissent de bons descripteurs?	

Traduction anglaise de ces mots-clé	

5.3. Autres éléments de réflexion ou d'actualité

5.3.1. L'existence du groupe « fouille de données »

Une question posée sur le forum ADBS-Infos sur les critères de pertinence des bases de données a révélé l'existence de ce groupe, comprenant pour l'instant le CRIN-CNRS, et l'INRIA Lorraine. La réponse complète de J. Ducloy figure en annexe 3. Au vu de l'absence de « résultats diffusables », un contact officiel me semble relever de la compétence du service IST.

5.3.2. La conférence TREC

Bien que cette conférence ne semble pas avoir d'intérêt direct pour l'utilisateur final de bases de données, il semble intéressant d'en souligner l'existence.

TREC, ou Text Retrieval Conference, est une conférence annuelle d'origine américaine dont la vocation est d'évaluer les systèmes automatisés de recherche d'information traitant en texte intégral de volumineux corpus. Depuis 1992, cette conférence affine sa méthode et étend les mesures d'évaluation qu'effectuent des participants issus du monde de la recherche et de l'industrie. Ces travaux sont essentiellement destinés aux concepteurs de ces systèmes d'accès à l'information textuelle (plus qu'à leurs utilisateurs), et vise à leur fournir des pistes de développement.

On peut lire à ce sujet un article complet de K. Lespinasse dans la revue Documentaliste, 1997, vol. 34, n°2.

5.3.3. Le problème Questel

Le serveur Questel a modifié ses tarifs à la hausse en 1997, apparemment sans consultation. Cela a pour conséquence que beaucoup de documentations annulent leur abonnement à ce serveur pour se tourner vers la concurrence lorsqu'elle offre le même service pour moins cher. C'est l'intention de plusieurs documentalistes du Cemagref, après s'être assurés qu'ils ont toujours accès aux mêmes bases via Dialog, ESA, etc...

En l'absence de réaction de la part de Questel, on ne peut que se poser des questions sur l'avenir à terme de ce serveur. C'est en tous cas ce qui ressortait de la lecture la liste ADBS-Infos au mois de juillet 1997. L'une de ces questions peut être : doit-on laisser disparaître sans réagir le seul serveur français, et augmenter ainsi l'hégémonie de produits anglo-saxons? Il semble que l'ADBS ait pris des contacts officiels avec Questel pour discuter du problème.

De plus la tendance des serveurs « géants » s'accroît avec l'annonce le 2 octobre dernier du rachat par Maid du groupe Knight Ridder Information (le producteur de Dialog) pour \$420 millions. « The Dialog Corporation » est le nom du nouveau groupe, qui propose ainsi plus de 900 bases de données représentant 6 milliards de pages de documents et 3 millions d'images.

5.4. La notion de « debriefing » et de retour d'expérience

5.4.1. Le debriefing

Les spécialistes en information insistent beaucoup sur cette notion. Dès que l'on interroge fréquemment les bases de données, il est nécessaire :

- d'une part d'en garder une trace, ce qui est indispensable pour des raisons de gestion,
- et surtout d'autre part d'avoir un retour sur l'usage qui a été fait de cette information par le demandeur. Cela s'avère particulièrement utile dans le cas de recherches sur profil ou de veilles, c'est-à-dire des recherches amenées à être mises à jour régulièrement : ce type de retour aide à améliorer le profil de recherche et à rester centré sur les préoccupations des chercheurs. D'une façon générale, cela permet d'évaluer la qualité d'une interrogation, son rapport bruit/silence, et donc d'améliorer les suivantes.

On préconise même, lorsque les interrogations atteignent une certaine masse, de constituer une petite base de données décrivant les éléments principaux de ces interrogations, de façon à les retrouver aisément.

Dans l'état actuel des choses au Cemagref, il n'est pas sûr qu'un produit de ce type soit un besoin réel. Par contre, dans l'éventualité d'une mise en commun d'expérience, ou d'une enquête nationale, une telle base pourrait être utile.

Dans cet ordre d'idées, voici également un projet de notice pour une base de données « debriefing » :

Projet de notice pour une bases de données « debriefing »

Date :	
Nom du chercheur :	
Nom du documentaliste :	
Groupement, Division :	
Département :	
Programme :	
Question en texte libre :	
Bases interrogées :	
Serveur :	
Descripteurs utilisés : (français/anglais)	
Nombre de références téléchargées :	
Temps passé :	
Coût :	
Nombre de références satisfaisantes	
Nombre de documents commandés	
Utilisation réelle des résultats (O/N)	
Si l'interrogation a été improductive, pourquoi ? Recherche des causes.	

Commentaires et souhaits éventuels	
------------------------------------	--

5.4.2. Propositions de mise en commun

Il ressort de ce travail, encore une fois, que l'idée d'une méthodologie toute faite applicable à toutes les situations d'interrogation a ses limites.

Il semble que ce qui peut manquer aux documentalistes du Cemagref sont des méthodes de travail liées à la pratique d'un outil, autant qu'une méthodologie.

Il leur manque parfois aussi, sans doute, un partage de ces méthodes.

Cette situation peut s'améliorer par la mise en commun d'expériences

- soit directement sur l'outil bases de données comme on l'a déjà suggéré;
- soit autour d'outils qui leur sont proches.

On peut illustrer ce propos en revenant à l'étude de cas.

Il avait été prévu, initialement, d'importer les références téléchargées sur le logiciel End Note Plus. Le temps qu'on pu consacrer les différentes équipes à ce travail ne l'ont rendu ni possible, ni nécessaire. De plus, le fait d'utiliser une procédure multibases rendait plus complexe l'élaboration d'un filtre sous End Note Plus. Les filtres prédéterminés existants dans le produit concernent tous des bases anglo-saxonnes; il n'y a pas, par exemple, de filtre pour la base Pascal.

Il semble intéressant, dans de telles circonstances, de faire remonter ce type de difficultés au sein du réseau des documentalistes. De même, le premier documentaliste à réaliser un filtre intéressant, comme celui pour Pascal, devrait en faire profiter ses collègues.

La structure pour ce type de mise en commun existe déjà; il s'agit des réunions de l'ensemble des documentalistes de l'établissement qui ont lieu environ trois fois par an. Le thème bases de données pourrait donc y être abordé soit de façon systématique, soit à chaque fois que cela est opportun pour y amener des éléments nouveaux.

Enfin, dans l'hypothèse où l'affichage d'un thème « bases de données » est susceptible de rebuter un certain nombre de documentalistes ou de chercheurs de l'Établissement, on pourra lui préférer l'affichage d'un partage d'expérience sur des outils qui leur sont proches. Il est de toutes façons nécessaire d'affiner l'usage de ces outils aussi en tant que tels. Voici une liste d'outils ou de thèmes qui pourraient convenir à cette approche:

- End Note Plus :
exemples d'utilisations du produits, filtres pour les bases les plus interrogées, etc...
- les outils de relais d'information vers les chercheurs:

Dans plusieurs groupements, les documentalistes rédigent des feuilles d'information sur des supports et avec des périodicités variés

D'une part ce type de vecteur est intéressant à partager, mais d'autre part c'est surtout son contenu qui est important de mettre à la disposition de tous.

- Le groupe « veille sur les sites Web »

Ce groupe me paraît être un excellent exemple de partage d'expérience. Le principe en est le travail en réseau : pour chaque thème, un responsable et des correspondants sont identifiés; ils s'emploient à repérer des sites Web intéressants, et à en mesurer l'intérêt réel en lien avec les scientifiques concernés. Enfin, les sites sont rassemblés par thème sous forme de liens hypertexte sur le serveur Web de l'Établissement.

La même méthode pourrait s'appliquer au repérage des bases pertinentes :

- identifier un responsable et des correspondants par thème;
- constituer des listes de bases de données par thème, enrichies au fur et à mesure des interrogations;
- partager cette information. La question peut se poser ici du vecteur idéal pour ce partage : doit-on privilégier le contact de type réunion, lui préférer la notion de groupe, ou bien encore utiliser le vecteur Internet ?

On pouvait se poser la question de la pertinence de ce vecteur en raison du fait que les documentalistes n'utilisent généralement pas les interfaces Web pour l'interrogation de bases de données en ligne, et leur préfèrent le mode « commandes » qui demeure plus efficace. On peut néanmoins penser que ce vecteur puisse être adapté pour au moins deux raisons :

- les interfaces Web sont de jour en jour plus puissantes et peuvent finir par remplacer le mode commandes (elles le font déjà pour bien des chercheurs qui interrogent eux-mêmes);
- la partie documentation des pages Web du Cemagref bénéficient actuellement de la dynamique du groupe de travail, et une rubrique supplémentaire sur les bases de données en ligne ne viendrait que renforcer le dispositif.

5.5. Conclusion

Ce travail qui s'est étalé sur deux années a permis de prendre conscience de la richesse des sources d'information disponibles dans notre Établissement, des moyens puissants dont nous disposons désormais pour y accéder, ainsi que pour accéder aux sources externes.

L'immensité de la ressource ainsi disponible nous fait parfois négliger les méthodes et outils adéquats.

Il s'est donc avéré particulièrement intéressant de faire à la fois un panorama des ressources et des outils disponibles.

L'existence d'un groupe professionnel (ou « transversal ») bien structuré et géré, celui des documentalistes du Cemagref, devrait permettre une bonne diffusion de ce type d'information.

On pourrait conclure sur le rôle que joue ce groupe de professionnels dans l'Établissement auprès de leurs utilisateurs pour amener ceux-ci à mieux maîtriser leurs propres ressources et à développer des démarches plus efficaces en matière de management de l'information. On peut sans risques affirmer qu'ils sont représentatifs des changements vécus par toute une profession ces dernières années, et citer à ce sujet une partie d'intervention du Prof. Jean Michel de l'École Nationale des Ponts et Chaussées sur le forum ADBS-info :

« Au premier degré, on trouve bien sûr ce que l'on appelle la formation des utilisateurs (user éducation, user instruction...) : savoir interroger une base de données, comprendre le fonctionnement d'une bibliothèque, etc... On peut aller plus loin et s'intéresser aux démarches de sensibilisation et formation (par des documentalistes) des divers acteurs de l'organisation en ce qui concerne les méthodes et technologies nouvelles (Internet, Intranet, etc.), la maîtrise de certains logiciels (concevoir des micro-bases de données personnelles) ou certains thèmes particuliers (la propriété industrielle, la propriété intellectuelle, les sources d'information européenne,...), etc... Un autre axe est celui du conseil ou du conseil-formation auprès des acteurs de l'entreprise ou de l'organisation ou de certains groupes d'acteurs : aider à mettre sur pied des outils personnalisés de traitement ou gestion de l'information, aider à réaliser des produits informationnels, etc...

Ces démarches professionnelles qu'on pourrait qualifier d'interventions de deuxième niveau sont importantes et le seront de plus en plus, les individus citoyens dans l'entreprise comme dans la société étant de plus en plus en mesure de se débrouiller par eux-mêmes pour trouver de l'information pertinente.

Ces démarches sont assez classiques pour les professionnels travaillant dans les établissements de formation, elles sont sans doute plus nouvelles dans d'autres milieux (entreprises, administrations, collectivités diverses). »

6. BIBLIOGRAPHIE

- Aït El Hadj S. - L'entreprise face à la veille technologique. Éditions d'Organisation, 1989
- Cemagref. - Bases de données bibliographiques utilisées au Cemagref, 1995
- Darrober M., Le Pottier N. - La recherche documentaire. Nathan, 1996
- Desval H., Dou H. (Dir) - La veille technologique. L'information scientifique, technique, et industrielle. Dunod, 1992
- Dialog Information Services. - Getting started on Dialog : a guide to searching. 1994
- Dou H. - Veille technologique et compétitivité. L'intelligence économique au service du développement industriel. Dunod, 1994
- Guillaumot R., Martinet B., Marti Y.M. - L'intelligence économique : les yeux et les oreilles de l'entreprise. Éditions d'Organisation, 1995
- Guinchat C., Menou M. - Sciences et techniques de l'information et de la documentation. Introduction générale. Les Presses de l'Unesco, 1981
- Jacobiak F. - Pratique de la veille technologique. Éditions d'Organisation, 1991
- Jacobiak F. - Exemples commentés de veille technologique. Éditions d'Organisation, 1992
- Jacso P. - Don't kiss boolean good-bye : it's AND not OR, let alone XOR. Information Today, Vol 11, n°2, pp. 22-24, 1994
- Lardy J.P. - Les accès électroniques a l'information : état de l'offre. ADBS Paris, 1993
- Lespinasse K. - TREC : une conférence pour l'évaluation des systèmes de recherche d'information. Documentaliste Sciences de l'Information, vol 34, n.2, pp. 74-81, 1997
- Martinet B., Ribault J.M. - La veille technologique, concurrentielle et commerciale. Éditions d'Organisation, 1989
- Ministère de l'Éducation Nationale, de l'Enseignement Supérieur et de la Recherche. - Des banques de données pour les étudiants, les enseignants, les chercheurs, 7ème édition, décembre 1995
- Polanco X. - L'infométrie, mode d'emploi. Éditions du CLPCF, Academia, 1992?
- Radasoa H.P. - Méthodes d'amélioration de la pertinence des réponses dans un système de bases de données textuelles. Thèse Doctorat Sciences De L'information, Université Paris 11, 1988
- Salton G., McGill M.J. - Introduction to modern information retrieval. McGraw -Hill International Editions, 1983, Singapore
- Such M.F., Perol D. - Initiation à la bibliographie scientifique. Promodis, Éditions du Cercle de la Librairie, 1987
- Tardieu H., Guthmann B. - Le triangle stratégique. Stratégie, structure et technologie de l'information. Éditions d'Organisation, 1991

7. ANNEXES

7.1. Annexe 1 : les opérateurs Booléens

7.2. Annexe 2 : extrait de la notice EndLink

7.3. Annexe 3 : courrier électronique sur le groupe fouille de données

Bonjour,

En réponse à Raymond Schiano :

>Subject: Demande d'infos : pertinence des bases de données

[...]

>Un travail collectif va être mené au Cemagref sur l'identification des banques de données les plus pertinentes pour l'Établissement.

[...] dans le domaine suivant : les critères pour définir la pertinence des bases par rapport à une thématique donnée, et tout élément de méthode pouvant s'y rapporter.

Le CRIN-CNRS, et l'INRIA Lorraine, laboratoires d'informatique, mettent en place une équipe sur "la fouille de données dans un contexte IST". L'identification des sources d'information pertinentes pour un établissement est donc au cœur de nos préoccupations. Cette équipe devrait regrouper plusieurs composantes, linguistiques (extraction d'éléments terminologiques d'un document), RFIA (pattern matching, classification), travaux sur les graphes (extraction de sous-graphes pertinents issus du Web notamment) et ingénierie du document structuré (SGML, TEI...).

Nous avons quelques idées sur la question, mais pas encore de "solutions diffusables". Nous espérons progresser rapidement dans ce sens.

Bien cordialement

Jacques DUCLOY

+-----+ | Jacques DUCLOY - ingénieur au
CNRS (INIST) |
+-----+ | Responsable de l'action "IST et
autoroutes de l'information" |
| pour l'INRIA Lorraine et le CRIN CNRS. |
+-----+ | Bâtiment LORIA | Tel. : (33) 83
59 30 38 |
Campus Scientifique,	(33) 07 83 65 87
BP 239	Fax. : (33) 83 27 83 19
F54506 Vandoeuvre les Nancy	Email : Jacques.Ducloy@loria.fr
+-----+