



**HAL**  
open science

# AUTOMATISATION DES RECHERCHES SUR INTERNET AU MOYEN D'AGENTS INTELLIGENTS

Marianna Mikolajczak

► **To cite this version:**

Marianna Mikolajczak. AUTOMATISATION DES RECHERCHES SUR INTERNET AU MOYEN D'AGENTS INTELLIGENTS. domain\_shs.info.docu. 2000. mem\_00000076

**HAL Id: mem\_00000076**

**[https://memsic.ccsd.cnrs.fr/mem\\_00000076](https://memsic.ccsd.cnrs.fr/mem_00000076)**

Submitted on 17 Feb 2004

**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.

*Université Henri Poincaré – Université Nancy II  
Institut National Polytechnique de Lorraine*

*PERNOD RICARD  
Centre de Recherche  
120, av. du Mal Foch  
94 015 Créteil CEDEX*

**RAPPORT DE STAGE DE D.E.S.S.  
Information Scientifique et Technique**

**AUTOMATISATION DES RECHERCHES SUR  
INTERNET  
AU MOYEN D'AGENTS INTELLIGENTS**

**Marianna Mikolajczak  
Promotion 1999-2000**

*Stage effectué au sein du Service de Documentation du  
Centre de Recherche **PERNOD RICARD**  
Mai- Septembre 2000*

**Responsable pédagogique :**  
M. Amos DAVID

**Responsable du stage :**  
Mme Jenny GROSSLERNER

## RÉSUMÉ

Internet représente pour un centre de recherche , une source d'informations non négligeable.

Les chercheurs du CRPR utilisent Internet à des degrés divers :

- consultations pour des recherches d'informations ponctuelles,
- consultations plus régulières pour une veille sur des sites stratégiques,
- messagerie électronique,
- listes de diffusion.

Les buts primordiaux à atteindre sont :

- le gain de temps d'une part,
- l'accroissement de la pertinence d'autre part.

De nombreux logiciels existent sur le marché et vont dans ce sens.

L'an passé, Sandrine Royer avait travaillé à la réalisation de grilles d'analyse d'agents intelligents de recherche d'information et de cartographie, et mis au point des tests sur certains d'entre eux.

Ce stage représente la suite logique de son étude : automatiser les recherches Internet au moyen de quelques uns de ces agents et poursuivre son travail en effectuant une veille sur les nouveaux produits.

Ce rapport vise à restituer la démarche utilisée pour la mise en place de l'automatisation des recherches sur Internet.

**Mots clés : Internet - Agent Intelligent - Web - Aspirateur de site - Métamoteur - Recherche documentaire - Automatisation - Agent d'alerte - Veille technologique -**

## SOMMAIRE

|   |      |
|---|------|
| Introduction  | p.2  |
| I. Présentation du groupe Pernod Ricard                                     | p.4  |
| A. La Holding Pernod Ricard   | p.5  |
| B. Le Centre de Recherche Pernod Ricard                                     | p.5  |
| II. Présentation du service de Documentation                                | p.8  |
| A. Présentation du personnel et de ses activités                            | p.9  |
| B. Présentation du matériel informatique et du réseau                       | p.10 |
| III. Présentation du sujet  | p.11 |
| III. L'automatisation des Recherches sur Internet                           | p.13 |
| A. Analyse du besoin  | p.14 |
| 1) Conclusions et bilan dressé en 1999                                      | p.14 |
| 2) Résumé des entretiens  | p.14 |
| B. Analyse des outils à disposition   | p.16 |
| 1) Exploration d'e-Catch  | p.16 |
| 2) Exploration de BullsEye, Copernic 99 et Strategic Finder                 | p.17 |
| 3) Exploration de Strategic Finder 1.5, Bullseye 2 Pro et Copernic 2000 Pro | p.19 |
| a) Strategic Finder 1.5   | p.19 |
| b) BullsEye 2 Pro   | p.20 |
| c) Copernic 2000 Pro  | p.21 |
| C. Méthodologie de recherche selon les différentes demandes.                | P.22 |
| 1) La surveillance de sites ciblés : Chimie et Conditionnement              | p.22 |
| 2) La surveillance de thèmes ciblés : Agronomie                             | p.23 |
| 3) Les "Profils Internet"   | p.24 |
| a) Récolte de l'information   | p.24 |
| b) Traitement de l'information  | p.28 |
| c) La diffusion de l'information  | p.28 |
| d) Le stockage de l'information   | p.29 |
| V. Veille sur les Agents Intelligents                                       | p.30 |
| Conclusion  | p.33 |
| Annexes   | p.34 |

### Internet ...

Internet a dernièrement franchi le cap du milliard de pages Web (source : France Inter – printemps 2000) ; il constitue aujourd’hui le “ réseau des réseaux ”, tant du point de vue technique que du point de vue humain : 200 millions de “ netsurfers ” s’y adonnent...

Par ses sites didactiques, ses bases de données et ses pages de “ News ”, il représente un média culturel, en fournissant librement l’accès à la connaissance ; par ses nombreux forums et listes de discussion (on en dénombre 70 000 !), il représente un foisonnement d’idées, un lieu d’échange et de discussion.

Si l’on considère le nombre de bases de données spécialisées accessibles sur le Web “ visible ” et “ invisible ”<sup>1</sup>, il s’avère être une source de données à part entière, aux côtés des banques de données documentaires classiques ; brevets, références d’articles (voire articles en texte intégral) sont en effet disponibles pour l’utilisateur averti.

Mais qu’on ne s’y trompe pas : la recherche d’informations sur INTERNET n’est pas simple. Face à un tel foisonnement d’informations il devient vital d’utiliser des outils adaptés pour faciliter sa recherche, et d’éviter les deux pièges d’Internet : le *manque de pertinence* et la *perte de temps*.

### Des agents intelligents...

Pour faciliter la tâche du butineur de toile, de nombreux outils existent. Les premiers à faire leur apparition étaient les annuaires, dont le but était de répertorier tous les sites existant sur la toile – tâche fastidieuse, et dont le résultat manque d’exhaustivité, face au nombre de sites créés chaque jour ... Les moteurs quant à eux , indexent inlassablement les pages Web au fil des visites automatiques effectuées par leur *spider*, et permettent de trouver l’information de façon plus fine, après saisie de mots-clés dans leur formulaires de recherche,... mais uniquement sur le faisceau de toile balayé par le moteur. Des métamoteurs sont alors apparus, qui, en interrogeant plusieurs moteurs en ligne, augmentent l’exhaustivité des recherches.

Mais comment conserver le résultat de ces recherches ? Comment classer, analyser (éliminer les doublons, les pages non pertinentes), diffuser les résultats intéressants ?

C’est dans ce cadre qu’entrent en scène les agents :

“ Un agent intelligent est une entité logicielle qui possède des attributs propres et qui agit dans le but d’accomplir un certain nombre de tâches au nom d’une autre entité (un autre agent ou une autre personne) ”<sup>2</sup> ; ces attributs sont :

- l’*autonomie*,
- la *collaboration*,
- l’*apprentissage* (s’adapte aux besoins en analysant ses actions passées)
- la *flexibilité* : il n’est pas entièrement prédéfini, autrement dit il est paramétrable,

---

<sup>1</sup> Voir glossaire en annexe

<sup>2</sup> Définition donnée par Carlo Revelli dans “ Intelligence Stratégique sur Internet ”, édition Dunod, préface de Joël de Rosnay.

- la *décision* du moment où il va entamer une action.

#### **Le Centre de Recherche Pernod-Ricard...**

Depuis 1995, le service de documentation du Centre de Recherche Pernod-Ricard (CRPR) s'intéresse de près à Internet : en 1996, des postes ont été installés pour les différents laboratoires ; en 1998, un guide d'utilisation d'Internet et un répertoire de sites Web ont été édités ; en 1999, un catalogue référençant les différents outils utilisables pour optimiser les recherches a été constitué et les résultats de l'évaluation de certains d'entre eux ont été diffusés – travail réalisé par Sandrine Royer, promotion 1998-99 du DESS I.S.T..

Comme une suite logique à ces différents travaux, le service de documentation souhaite automatiser les recherches sur Internet, au moyen des différents agents et outils testés, et de poursuivre une veille technologique sur les différents outils disponibles sur le marché.

Ce rapport décrit la méthodologie mise en œuvre, et se décline en deux parties :

- une première phase correspondant à une étude du besoin, une prise de connaissance des différents outils disponibles et une sélection des outils utilisables,
- une deuxième phase correspondant à la mise en place concrète de ces recherches et leur diffusion sur le réseau.

\*\*\*

I. PRESENTATION DU GROUPE  
PERNOD RICARD

## **A. La holding Pernod Ricard**

Le groupe Pernod Ricard est aujourd'hui le premier producteur de vins et spiritueux dans la zone Euro, et le cinquième au plan mondial avec un chiffre d'affaires qui atteint les 20,5 milliards de francs.

L'activité du groupe s'organise de fait en trois pôles :

- les vins et spiritueux - dont les fameux anisés,
- les préparations et boissons aux fruits,
- la distribution.

Implantée sur les cinq continents, la holding Pernod Ricard est avant tout une structure décentralisée, dont les principales sociétés de production sont : *Irish Distillers, Ricard, Pernod, CSR-Pampryl, Campbell distillers, Orlando Wyndham, Austin Nichols*. Toutes gèrent leur propre réseau de filiales tout en étant rattachées au groupe.

Dans le secteur des préparations aux fruits, *Sias-MPA*, leader mondial du secteur, dispose de filiales et de centres d'élaboration sur tous les continents.

Une liste exhaustive des différentes marques du groupe se trouve en annexe.

## **B. Le Centre de Recherche Pernod Ricard.**

Tous les produits du groupe Pernod Ricard sont élaborés à partir de matières premières agricoles ; le groupe sélectionne et transforme chaque année plus de 2,1 millions de tonnes de fruits, de plantes et de céréales.

Pernod Ricard intervient sur tous les maillons de la chaîne d'élaboration : amélioration et sélection des matières agricoles, process d'élaboration, distillation, conditionnement. Il consacre de ce fait des efforts importants à la recherche et au développement.

Le Centre de Recherche compte sept laboratoires spécialisés, qui participent activement à la recherche fondamentale dans le secteur agro-alimentaire. Ces laboratoires sont les suivants :

### **▪ Agronomie :**

Ses différentes unités ont une triple mission :

- posséder les meilleures connaissances sur les techniques culturales des fruits et plantes aromatiques utilisés dans la fabrication des boissons et préparations de fruits ;
- rechercher et mettre au point de nouvelles techniques de production agricole mais aussi créer de nouvelles variétés ;
- conseiller et assister les agriculteurs qui fournissent au groupe Pernod Ricard les fruits et plantes dont il a besoin.

▪ **Chimie :**

Il existe deux pôles de chimie, dont les activités sont séparées :

- Un pôle où les analyses physico-chimiques qui y sont effectuées permettent de déterminer la composition des arômes et de pouvoir, par la suite, mieux exploiter le potentiel aromatique des fruits et des plantes. Ce laboratoire est ainsi en mesure de repérer une molécule aromatique de synthèse et de la distinguer de son équivalent d'origine naturelle. Ceci revêt un enjeu important car seules ces dernières ont le droit à l'appellation " arôme naturel ".
- Un pôle effectuant des contrôles d'authenticité, par le biais d'analyses isotopiques. Il est ainsi possible de détecter si un jus de fruits est entier ou s'il a subi un apport de sucre. A ce titre, le laboratoire tient la place d'un véritable " service de répression " des fraudes .

▪ **Nutrition et sécurité :**

Tous les produits doivent répondre aux exigences de sécurité pour les consommateurs, et contribuer à leur équilibre nutritionnel. Cette double priorité s'appuie sur une bonne connaissance de la réglementation sur tous les constituants des boissons et un dialogue permanent avec les laboratoires des filiales et du centre de recherche, de façon à prendre en compte leurs contraintes et les conseiller pour garantir la conformité des produits.

▪ **Analyse sensorielle et nouveaux produits :**

Le rôle du laboratoire de formulation et d'analyse sensorielle touche à un aspect primordial de la conception d'un produit : l'évaluation de ses propriétés organoleptiques, la mesure de ses qualités sensorielles, et bien sûr, son appréciation par le consommateur. Goût, couleur, odeur, consistance, texture, tenue en bouche : autant de paramètres testés par les spécialistes et dégustateurs du laboratoire, véritables garants de la qualité du groupe.

▪ **Conditionnement :**

Ultime maillon du centre de recherche, le laboratoire de conditionnement et d'études industrielles a trois missions :

- l'étude des matériaux de conditionnement et d'emballages, et plus particulièrement, ceux destinés au contact alimentaire, qui implique une sélection auprès des fournisseurs de matières premières d'emballage : verre, alliages ou plastique.
- le développement d'emballages nouveaux, en collaboration avec les services marketing sur les innovations en matière de " packaging ".
- l'étude de nouveaux procédés de conditionnement (procédés ultra-propres, par exemple) et leur mise en œuvre.

▪ **Technologie :**

Les principaux procédés de transformation nécessaires à l'élaboration des boissons et préparations de fruits sont la distillation, la fermentation, l'extraction, la filtration, la concentration, la pasteurisation ou la stérilisation.

La maîtrise parfaite de ces procédés ainsi que des matériels est la tâche du laboratoire de technologie qui, pour tester de nouvelles techniques et les mettre au point, dispose d'un atelier pilote, reproduisant à petite échelle tous les procédés mis en œuvre dans les unités du groupe.

▪ **Pharmacologie :**

L'importance de la connaissance des mécanismes qui conduisent à la dépendance alcoolique a justifié la création d'un laboratoire de pharmacologie entièrement consacré à cette étude.

Ses travaux ont ainsi montré comment une consommation excessive de boissons alcoolisées pouvait, à long terme, modifier la structure et la composition des membranes cellulaires, en particulier celles du cerveau, et donc le comportement de l'individu. Cette recherche pourrait permettre de comprendre pourquoi, alors que la grande majorité des consommateurs restent des buveurs modérés, un petit nombre d'entre eux perdent la maîtrise de leur consommation. Le laboratoire a donc mis au point un modèle de dépendance envers l'alcool pour en percer plus facilement les mécanismes biologiques.

▪ **Microbiologie :**

Organismes unicellulaires vivants, les levures interviennent directement dans l'élaboration des boissons alcoolisées : ce sont elles qui transforment les sucres contenus dans les matières premières (céréales, jus de fruits) en alcool. Par leurs activités métaboliques, elles produisent un certain nombre de composés aromatiques qui contribuent à la qualité organoleptique des boissons.

Le laboratoire de microbiologie consacre l'essentiel de son activité à la connaissance et à l'amélioration génétique des performances de ces micro-organismes en créant de nouvelles races ou variétés, mieux adaptées à l'élaboration de boissons nouvelles.

## **II. PRESENTATION DU SERVICE DE DOCUMENTATION**

## A. Présentation du personnel et de ses activités.

Le service de documentation est aujourd'hui composé de 3 personnes :

- Mme Jenny Grosslerner, documentaliste de formation scientifique et responsable du service.
- Melle Cécile Kéruzec, chargée de la gestion des abonnements et des profils bibliographiques.
- Melle Cécile Fournier, chargée de la gestion des commandes.

L'activité du service peut être résumée selon quatre grands axes :

- La recherche documentaire qui est effectuée de deux façons :
  - ponctuellement en réponse à des demandes précises,
  - automatiquement selon des profils bibliographiques établis et de chronicité variable : 2, 4 et 6 mois.Ces derniers sont lancés de façon automatique sur les serveurs Dialog et STN. Le logiciel de communication utilisé, Infolog d'IUK, est un logiciel de communication permettant d'accéder et d'interroger les bases de données des serveurs en ligne et d'utiliser un courrier électronique via une ligne téléphonique, ou une session Telnet.  
Pour le reformatage des données, le logiciel utilisé est Infotrans, qui permet l'intégration des notices bibliographiques sur la base de données interne.
- La gestion des abonnements : le service est abonné à 140 revues et à une dizaine de CD-ROM. Cette gestion comprend le dépouillement du courrier, bulletinnage, les réclamations aux éditeurs (RowCom et Swets), la numérisation et l'océrisation des sommaires de revue en vue de l'alimentation du logiciel documentaire (Basis), les réabonnements, la diffusion.
- La gestion des commandes : les articles scientifiques, les brevets, les normes et les ouvrages sont commandés auprès d'éditeurs attitrés - la British Library, l'INIST, les éditions Lavoisier - ainsi qu'auprès de grands organismes spécialisés tels que "Leatherhead - Science Technology and Information for Food Industry" ou "Campden - Food and Drink Research Association". Cette activité regroupe également la gestion des prêts (thèses, mémoires, ouvrages) auprès de l'I.N.T.D., de la British Library ou auprès d'universités de façon ponctuelle. Ces commandes deviennent des références dans la base de données interne et sont accessibles à tout le personnel.
- Une activité Internet, attribuée depuis quelques années aux stagiaires. Elle s'inscrit dans une mission constante : accroître l'autonomie de l'utilisateur, en développant l'accès aux bases internes, au réseau de CD-ROM, à Internet.

## B. Présentation du matériel informatique et du réseau.

Le service dispose de :

- 2 serveurs SUN, l'un utilisé pour Basis et Techlib, l'autre pour la ligne Transpac,
- 1 serveur de type "Novel Netware", qui partage les applications sur CD-ROM : il est utilisé pour la gestion du réseau de CD-ROM
- 2 serveurs NT :
  - 1 contrôleur principal de gestion qui gère le réseau NT pour le centre de recherche, la messagerie interne (CC-Mail), les sauvegardes, les imprimantes.
  - un contrôleur secondaire de domaine, qui seconde le contrôleur principal et sert de serveur Proxy pour la gestion des pages HTML dans Techlib.
- 35 postes réseau et autant hors réseau, dont 7 postes dédiés à la documentation plus 1 à l'administrateur réseau.
- 10 postes Internet :
  - 1 au laboratoire de technologie
  - 1 à celui de chimie du bloc 3
  - 1 à celui de chimie du bloc 5
  - 3 en documentation (dont un au bloc 3)
  - 1 au laboratoire de Microbiologie
  - 1 à celui d'Analyse Sensorielle
  - 1 en pharmacologie
- 2 postes sécurisés servant à la messagerie Pernod Ricard (Lotus Notes).

\*\*\*

### **III. PRESENTATION DU SUJET**

“ L’automatisation des recherches sur Internet au moyen des agents intelligents et des outils de cartographie du service de documentation ” s’inscrit dans la continuité du travail réalisé par les stagiaires précédents :

- accès à Internet et élaboration d’un répertoire d’adresses et de sites Web pour les utilisateurs du CRPR (1998),
- évaluation des agents intelligents de recherche d’information et des logiciels de cartographie (1999).

Le but de ce stage est de rendre les recherches sur Internet les moins voraces en temps, tout en leur conservant une pertinence susceptible de les rendre exploitables, comme source d’information scientifique à part entière.

Ce stage s’est décliné en plusieurs axes :

1°) la mise à jour du répertoire Internet pour les différents services et filiales.

2°) Une étude de besoins auprès des différents laboratoires, donnant lieu à l’automatisation des recherches à proprement parler. Celle-ci a été séparée en deux phases distinctes :

- des tests et un paramétrage des agents intelligents,
- un accès des rapports de recherche et profils Internet sur le réseau.

3°) Une veille technologique sur les différents agents intelligents et logiciels utilisables sur Internet.

#### **IV. L'AUTOMATISATION DES RECHERCHES SUR INTERNET**

## A. Analyse du besoin pour chaque laboratoire

### 1. Conclusions et bilan dressé en 1999.

Au départ de Sandrine Royer en septembre 1999, un travail considérable avait été accompli : non seulement un catalogue exhaustif des différents outils et logiciels Internet avait été réalisé, mais aussi une évaluation de recherches auprès de deux laboratoires en utilisant les outils les plus performants, et qui se sont avérés être aussi les plus plébiscités par les chercheurs.

Après une prise de connaissance des différents supports d'informations concernant ces outils – livres, articles, et bien sûr rapports -, je suis allée à la rencontre des principaux intéressés : les responsables des différents laboratoires.

### 2. Résumé des entretiens.

Ces rencontres, qui se sont déroulées sur un mois, avaient plusieurs buts :

- me mettre au courant de leurs activités respectives, afin de mieux cibler leurs domaines de recherche,
- prendre connaissance de leurs attentes en matière d'automatisation de recherches et définir avec eux leurs sujets d'information,
- m'informer de l'utilisation qu'ils avaient d'Internet et des outils à disposition dans le service, après deux ans de mise en place de l'activité Internet,
- une présentation de ces outils et un rappel de leurs fonctionnalités (pour ceux qui ne les utilisaient pas),
- me faire connaître.

Les sujets abordés lors de l'entretien se devaient d'être homogènes entre les différents intéressés – la suite de l'étude en était déterminante !

C'est pourquoi les mêmes points étaient abordés lors de l'entrevue :

- leurs axes et thèmes de recherche,
- leur utilisation d'Internet : plutôt de la recherche d'informations ? plutôt un suivi de sites ? sont-ils abonnés à une liste de diffusion dans leur domaine ?
- leur connaissance des différents outils à disposition dans le service et leur utilisation, ce qu'ils en attendaient, quels étaient leurs points forts, leurs points faibles selon eux,
- leurs souhaits en matière d'automatisation : un système d'alerte, un envoi de dossiers par e-mail ; et s'ils ne savaient pas, la description de l'outil idéal.

Une synthèse a été effectuée à partir de ces entretiens, sous la forme d'un rapport.

Les résultats sont donnés sous la forme d'un tableau ci-après :

| <i>Laboratoire</i> | <i>Demande exprimée</i>  |  |   |
|--------------------|--|--|---|
|                    | <i>Automatisation</i>  | <i>Recherche</i>   | <i>Autre</i>                                    |
| Technologie        | <i>Contamination par "Alicyclobacillus acidoterrestris"</i>  |  |   |
| Nutrition-sécurité | <i>Orange, pomme, myrtille.. sous un aspect nutritionnel</i>   |  | Réorganisation des signets du service.          |
| Agronomie          | <i>Polluants organiques persistants et pesticides dans différents pays</i>   |  |   |
| Chimie bloc 3      |  |  | Veille sur le matériel chimique et les congrès. |
| Pharmacologie      | -  | -  | -   |
| Conditionnement    | <i>Polymères barrière</i>  |  |   |
| Nouveaux produits  |  | Recherche ponctuelle sur les études en matière d'habitudes alimentaires et préférences nutritionnelles, en France et à l'étranger. |   |
| Chimie bloc 5      | <ul style="list-style-type: none"> <li>- Métaux lourds dans les jus de fruit et les boissons</li> <li>- Mycotoxines : limites de tolérance</li> <li>- Glycérol et acide gluconique : méthodes d'analyse, taux naturels existant dans le jus de raisin</li> </ul> |  |   |
| Microbiologie      | ?  |  |   |

**Tableau 1. Récapitulatif des demandes émanant des différents laboratoires.**

Cette synthèse a fait apparaître deux types de demandes, si l'on exclut les travaux non directement liés au sujet (recherche ponctuelle pour le laboratoire d'analyse sensorielle et réorganisation des signets du laboratoire de nutrition) :

- d'une part, une demande de recherches automatisées classiques, que l'on pourrait assimiler à des " profils-Internet ", et correspondant de fait à une veille qu'effectuent les laboratoires concernés de façon permanente,
- d'autre part, une demande s'apparentant plutôt à la surveillance d'une information d'ordre scientifique et législatif à la fois, et décisive dans l'orientation de ces activités : celle du laboratoire d'agronomie.

J'énoncerai dans le chapitre consacré à la méthodologie, la stratégie de recherche qui a été adoptée pour ces différentes demandes, et qui a été élaborée en parallèle de l'étude des outils à disposition.

## **B. Analyse des outils à disposition.**

Le catalogue établi par Sandrine Royer l'an passé et qui visait à répertorier les différents logiciels/outils/agents du marché, dénombrait :

- 8 aspirateurs de sites web
- 4 méta-moteurs évolués
- 16 " super " méta-moteurs
- 12 logiciels de veille, de partage de connaissances et d'analyse de texte.

Tous, bien sûr, n'avaient pu être testés – beaucoup de logiciels peuvent être téléchargés gratuitement, mais certains, aux fonctionnalités intéressantes et souvent chers, ne le sont pas. Par ailleurs, leur évaluation peut prendre beaucoup de temps.

A mon arrivée dans le service, mon poste était doté, comme les deux autres postes réservés à la documentation, des logiciels suivants :

- Aspirateurs :
  - e-Catch
  - Memo Web
- Super méta-moteurs :
  - Copernic 99
  - Strategic Finder 1.3
  - Bulls Eye
  - WebSeeker
- Logiciels de cartographie :
  - U-Map Universal 1.8b

La première étape de mon stage a été une familiarisation avec les différents outils présents sur le poste.

### 1) Exploration d'e-Catch

Outre sa fonction principale d'aspirateur de site, ce logiciel, dont la presse spécialisée a donné et donne toujours des échos très favorables, permet :

- d'indexer des résultats en texte intégral et donc de suivre dynamiquement un site,

- de mettre à jour des sites à la fréquence de son choix.

Des tests ont été effectués sur des pages dont le contenu est constamment modifié (site des quotidiens “ Le Monde ” et “ Libération ” en particulier), et des aspirations ont été lancées pour ces sites.

Malheureusement, il s’est avéré présenter un inconvénient de taille : *il n’est pas programmable*. Le rafraîchissement des pages n’était possible que dans la mesure où le logiciel est mis en route manuellement. Or, la surveillance de sites ou pages telle qu’elle avait été envisagée n’est possible qu’avec un système d’alerte en temps réel, et qui fonctionne si les rafraîchissements de pages sont effectués régulièrement.

Par conséquent, son utilisation pour la surveillance de sites n’a pas été retenue.

## 2. Exploration de BullsEye, de Copernic 99 et de Strategic Finder.

BullsEye, Copernic 99 et Strategic Finder sont tous deux des “ super ” méta-moteurs de recherche, c’est-à-dire qu’ils offrent, en plus de la possibilité d’interroger un nombre important de moteurs (fonction “ méta-moteur ”), des fonctions d’analyse de résultats (taux de pertinence, dédoublement, exclusions des liens périmés, raffinage et indexation des résultats) de même qu’une recherche de type avancé (opérateurs booléens ET, OU, SAUF et ( ) pour Copernic, les mêmes plus le NEAR pour BullsEye).

Des requêtes ont été testées pour avoir une idée de leurs performances.

Pour BullsEye et Copernic, elles se sont avérées d’emblée intéressantes au niveau de la vitesse d’obtention des résultats, de même que des points de vue de convivialité et de facilité d’utilisation. Ce furent celles correspondant à la demande du service d’agronomie, et qui concernaient les pesticides.

Pour Strategic Finder, les résultats furent plus mitigés au niveau de la vitesse et au niveau de la pertinence, mais contrebalancés par une fonctionnalité intéressante n’existait pas pour les deux précédents : la possibilité d’ajouter des sources (moteurs, sites) “ stratégiquement ” choisis, afin d’optimiser la pertinence *avant* de lancer la recherche – pour les deux autres, le balayage effectué par les moteurs interrogés ne donne de la pertinence qu’*après* le “ raffinage ” (Copernic) ou l’“ analyse ” (BullsEye).

La lecture de la presse spécialisée et la visite des sites des différents éditeurs m’a permis de me tenir informée des différentes mises à jour disponibles pour chacun de ces agents, qui présentaient par ailleurs des fonctionnalités supplémentaires (voir fiches outils).

C’est ainsi que Copernic 2000, BullsEye 2.5 et Strategic Finder 1.5 ont été téléchargés sur les 3 postes de la documentation.

Voici sous la forme d’un tableau, résumées ces différentes fonctionnalités :

Voir tableau en annexes

### 3) Exploration de Strategic Finder 1.5, BullsEye 2 Pro, et Copernic 2000 Pro.

#### a) Strategic Finder 1.5 (voir fiche-outil en annexe)

Cette version a été téléchargée sur le Web. Le logiciel ne prévoyait pas de sauvegarde des anciennes recherches...

La force de cet outil réside dans la possibilité d'interroger des "plugg-in" spécifiques (les "plugg-in" sont des fichiers permettant l'interrogation de plusieurs dizaines de moteurs de recherche, ou de sites concernant un domaine précis).

Il existe des "plugg-in" gratuits (ceux que l'on obtient lorsque l'on télécharge la version gratuite) et des "plugg-in" payants, qui sont constitués de plusieurs dizaines de sources spécifiques à un domaine ; la société Digimind édite des versions spécialisées dans des domaines très pointus : transports, biotechnologie et surtout, le plugg-in "AgroFinder" dont les sources sont le plus à même d'intéresser le CRPR.

Strategic Finder offre également la possibilité de construire soi-même des "plugg-in", en intégrant dans des fichiers spécifiques les sites, moteurs et pages de son choix.

L'idée de départ a donc été de construire des "plugg-in" pour chacun des laboratoires afin que l'interrogation du web soit ciblée. Cette possibilité était en fait surtout envisageable pour le sujet d'agronomie.

Des tests ont été réalisés afin d'explorer ces fonctions poussées.

#### ▪ Intégration de moteurs dans le "plugg-in"

Il s'agit bien sûr des moteurs internes à des sites.

Pour ce faire, il faut, à partir du moteur de son choix :

- lancer une recherche,
- consulter la 2<sup>ème</sup> ou 3<sup>ème</sup> page de résultats,
- copier dans l'espace de la fenêtre réservé à la requête du moteur (dans le plugg-in) la syntaxe exacte rapportée dans la fenêtre du navigateur réservée à l'adresse (répertoire CGI du moteur),
- apporter les modifications suivantes : remplacer le mot-clé utilisé dans la requête par <keyword>, remplacer le chiffre correspondant à l'index de pages par <index>>,
- paramétrer l'interrogation en fonction des booléens que le moteur choisi utilise : ET, + ou l'espace vide correspondant au ET implicite.
- paramétrer l'index de page selon qu'il est basé sur le résultat ou le numéro de page, et selon qu'il est basé sur 0 ou sur 1 (Exemple : dans le cas d'un index basé sur les résultats, si le moteur rapporte 10 résultats par page et le 1<sup>er</sup> résultat de la 2<sup>ème</sup> page est 11, c'est que l'index est basé sur 1, si c'est 10, c'est qu'il est basé sur 0).

Les résultats d'intégration de moteurs étaient décevants : ce n'étaient parfois pas les moteurs les plus intéressants qui fonctionnaient au moment de l'interrogation...

- Intégration de pages.

La marche à suivre était relativement simple, puisqu'elle consistait à recopier l'adresse en “.html ”.

- Intégration de sites.

Il suffisait de copier l'adresse “ www.site.com ”

Le choix des sources, déterminant pour la réalisation du “ plugg-in ” d'agronomie en particulier, a été réalisé à partir :

- des adresses correspondant à leurs signets, et qui méritaient d'être exploitées,
- la bibliographie-Internet d'un étudiant d'agronomie qui avait travaillé sur ce sujet,
- le livre “ Veille technologique et Stratégique sur Internet ”, qui délivrait des adresses intéressantes dans ce domaine.

Le choix des pages et des sites d'une façon générale était fait en fonction de :

- sa crédibilité (sites officiels et incontournables),
- le degré de popularité, c'est-à-dire si beaucoup de petits sites ou associations pointaient des liens vers ces pages ou ces sites,
- la possibilité d'aboutir à du texte et surtout, de l'information en lien avec le sujet.

Les résultats de ces essais sur Strategic Finder s'avèrent mitigés au moment de l'interrogation du moteur pour ces plugg-in “ construits ”. En effet, Strategic Finder avait montré des faiblesses au moment de l'interrogation sur le Web –il “ plantait ” souvent au moment du téléchargement des pages trouvées. Un “ plugg-in ” construit étant moins important en taille, il était justifié d'espérer un meilleur résultat. Or, il n'en a rien été.

Strategic Finder, au jour d'aujourd'hui, continue de “ planter ”, mais la possibilité d'obtenir une démonstration d'“ AgroFinder ” est envisagée (il n'existe pas de version d'évaluation), et surtout, la perspective d'une nouvelle version plus puissante à la rentrée nous a incité à laisser ce produit de côté pour le moment.

#### b) BullsEye 2 Pro.(voir fiche-outil en annexe)

Cette nouvelle version de BullsEye se différencie de la précédente par un nombre considérable de sources supplémentaires : plus de 800 contre 450 dans la précédente. Il a donc été jugé utile d'acquérir cette nouvelle version ; par ailleurs, que cette version offrait la possibilité d'une période d'essai de 30 jours gratuite.

Des difficultés sont apparues lors de son l'installation :

1- Dans un premier temps, l'impossibilité de récupérer les anciens rapports de recherche de l'ancienne version.

Ceci nous a conduit à désinstaller puis à réinstaller BullsEye. Cette réinstallation, nous a permis la récupération des anciens rapports, mais nous prive de la possibilité de les consulter directement à partir du logiciel : il faut passer par l'explorateur...

2- Par la suite, il s'est avéré que BullsEye 2 Pro ne disposait pas de sa fonctionnalité première, à savoir *l'automatisation des recherches* : aucune recherche ne pouvait être lancée automatiquement à partir de l'agent de veille (" Tracker "), de même qu'aucune surveillance de page.

De nombreux e-mails ont été envoyés au service technique aux U.S.A. (sans réponse), de même que des appels téléphoniques, qui ne nous ont pas permis d'obtenir la réponse à ce problème.

Les problèmes rencontrés avec BullsEye nous ont amenés à nous questionner sur la valeur de ce produit dans une perspective de long terme :

- peut-on se fier sur un produit ne prévoyant pas de sauvegarde des anciens dossiers lors de sa mise à jour ?
- peut-on se passer d'une automatisation des recherches - c'est-à-dire lancer des recherches à la main - , sachant le nombre important de recherches (et donc, sous-entendu le nombre de requêtes) à poser pour chacun des demandeurs ?

La réponse à ces questions a été fournie lors de l'exploration de Copernic 2000 Pro.

#### c) Copernic 2000 Pro.(voir fiche-outil en annexe)

A l'inverse de BullsEye Pro, Copernic 2000 Pro prévoit, lors de son installation, au moment de l'exécution du module téléchargé sur le disque dur, *la sauvegarde des anciennes recherches effectuées*.

Au milieu de ses nombreuses fonctionnalités, il en figure une qui n'est pas à négliger : la possibilité d'interroger le Web en français. Cette caractéristique qui ne figure sur BullsEye peut s'avérer très intéressante pour certaines recherches ayant trait à des données législatives, qui pour l'essentiel nécessitent un balayage du Web national, voire européen (Cas des laboratoires de Chimie 5 et de Conditionnement).

L'exploration du logiciel n'a pas fait apparaître de problèmes notables ; des recherches automatisées fonctionnent sur Copernic et donnent à ce jour des résultats satisfaisants.

\*\*\*

Cette étape, quoique longue, a permis l'exploration fine de chaque logiciel répertorié.

Leur mise à jour a permis d'affronter des problèmes qui n'avaient pas été rencontrés l'an passé, ce qui permet d'avoir une vision de leurs fonctionnalités à long terme, essentielle dans l'expectative d'une automatisation pérennisable des " profils – Internet ".

## C. Méthodologie de recherche selon les différentes demandes.

Comme nous l'avons déjà vu, les sujets de recherche proposés par les chercheurs sont très différents les uns des autres, et ce pour diverses raisons :

1° du fait de leur contenu :

Les domaines d'études des différents chercheurs sont loin de se ressembler, même s'ils travaillent a priori sur les mêmes axes de recherche.

Ainsi, pour un sujet concernant la fraise, le laboratoire de nutrition s'intéressera à ses polyphénols (communément appelés "tannins"), et en particulier à l'impact de ceux-ci sur la santé ; celui d'agronomie s'intéressera aux pesticides autorisés sur les cultures de fraises ; le laboratoire d'analyse sensorielle, enfin, s'intéressera à ses qualités gustatives...

2° du fait des sources de données qu'ils ont l'habitude d'utiliser,

3° *Last But Not Least*, du regard qu'ils portent sur Internet. Des recherches sur Internet fructueuses incitent plus à la confiance que des recherches ou trop bruyantes, ou stériles, etc...

A ces demandes diverses correspondent des outils aux fonctionnalités différentes, qui seront jugés plus ou moins intéressants selon les demandeurs, selon qu'il a une bonne connaissance d'Internet et une utilisation régulière des outils ou non.

Cette partie a pour but de décrire la méthodologie employée selon les deux types de demandes qui se sont dégagées à la suite des entretiens menés :

- les profils-Internet : Chimie 5, Nutrition, Technologie, Conditionnement.
- la surveillance de sites / thèmes en rapport avec des sujets ciblés : Agronomie, Conditionnement, Chimie 3.

Une demande de recherche Internet ponctuelle a été demandée, qui ne sera pas traitée, ne cadrant pas directement avec le sujet (Laboratoire d'analyses sensorielles), de même que la réorganisation du "bookmark" du service de nutrition.

### 1. La surveillance de sites ciblés : Chimie 3 et Conditionnement

#### ▪ Chimie 3 :

Ce laboratoire a émis la demande d'une surveillance des sites fournisseurs de matériel d'analyse, de même que la surveillance de congrès en chimie.

#### ▪ Conditionnement.

Ce laboratoire, en plus d'une recherche de type profil, souhaitait être renseigné sur les nouveautés apparaissant sur les sites de certains syndicats professionnels dans le domaine des plastiques, et dont la liste a été fournie.

Ces deux laboratoires répondent à un même besoin : les sites sont dans les deux cas clairement définis, la requête ne correspond aucunement à une recherche d'informations.

La solution la plus adaptée revêt la forme d'un système d'alerte.

Un système d'alerte, d'une façon générale et à ce jour, peut être :

- soit un agent d'alerte (technologie "pull"\*), où l'on reçoit l'information sous la forme d'une alerte par e-mail (exemple : "Mind-It" de Netmind)
- soit un agent d'alerte off-line (logiciel client que l'on installe) : c'est le cas des aspirateurs de site de deuxième génération, qui permettent d'être alertés au moment du rafraîchissement des pages téléchargées, par un e-mail ou un message apparaissant à l'écran, dès qu'une modification sur la page ou le site est détectée.

Les systèmes d'alerte en ligne - type "Mind-It" - peuvent être, dans l'absolu, une source de danger au niveau de la confidentialité. (source : Confidentialité sur Internet - rapport interne d'une conférence IDT et datant du 20/06/00)

L'installation d'un aspirateur peut-être une solution adaptée, dans le cas où la surveillance est régulière.

Ce type d'outils nécessite un paramétrage relativement simple, où sont entrées les adresses des sites ou pages à surveiller, suivant un calendrier de rafraîchissement des aspirations, de même que les éléments à surveiller sous la forme de mots-clés : l'alerte n'est lancée que si des modifications en lien avec les mots clés apparaissent sur la page.

L'étude de Webspector (Illumix) est en cours. Cet aspirateur de 2<sup>ème</sup> génération (il autorise une alerte ciblée) est automatique, contrairement à E-Catch (La Mine). Il semble aussi plus simple d'utilisation, et plus convivial.

## 2. La surveillance de dossiers d'étude ciblés : Agronomie.

La demande du laboratoire d'agronomie était la suivante :

Certains pays utilisent des pesticides qui ne sont pas ou plus autorisés en France, et certains mêmes des pays souffrant de la pollution par des POP (Persistant Organic Polluant) ou des métaux lourds. Il s'agissait de trouver :

- les cartes de contamination de ces différents polluants,
- les quantités émises par an et par pays,
- la législation (limites par produit et par pays), dans le monde entier.

L'idée pour ce sujet a été de reprendre les données d'un stagiaire en Agronomie, afin d'utiliser les sites trouvés, et les inclure dans un "plugg-in" construit à cet effet : le demandeur n'aurait plus eu qu'à interroger de façon ciblée les sites pertinents en utilisant des mots-clés (pesticide, "POP") et suivre ainsi l'évolution de la législation et des données sur la pollution.

Des sites tels que ceux de la FDA (Food and Drug Administration) et de la FAO - délivrant en particulier des rapports de commission du Codex Alimentarius - furent intégrés au "plugg-in" ainsi que d'autres sites en lien avec la législation des pesticides. (voir plugg-in en annexe)

---

\* voir Glossaire

Cette solution s'avère intéressante, mais nécessite de façon complémentaire une recherche préalable des sites pouvant fournir des informations pertinentes : un plan de monitoring contrôlant les pesticides en Europe n'existe pas forcément en Afrique du Nord ou au Mexique, ou au Chili.

Elle dépasse le cadre d'une simple automatisation car elle nécessite le regard averti du spécialiste...

### 3. Les "Profils-Internet"

Ces profils représentent le pendant des profils bibliographiques tels qu'ils sont effectués dans le service : automatisés, ils tournent régulièrement sans nécessité de renouveler les requêtes. Ils correspondent à des sujets d'étude permanents du CRPR.

Des recherches automatiques sur Internet devaient, pour leur mise en place répondre à certaines exigences :

- la fiabilité du support au sens technique : pas de " bugg " - ou le moins possible..
- l'automatisation des recherches à proprement parler, soit la possibilité de paramétrer l'heure, le jour, le mois selon un calendrier établi en accord avec le demandeur.
- la facilité de gestion. Après mon départ, ces recherches automatiques continueront à " tourner ", elles devront être gérées au niveau du traitement des résultats, et être diffusées. Cette gestion se doit d'être simple afin de ne pas surcharger la personne qui en aura la responsabilité.
- le stockage des résultats, afin que les résultats de recherche puissent rester accessibles dans le temps.

Ces profils ont été à ce jour définis pour 4 laboratoires :

- conditionnement
- nutrition
- chimie
- technologie.

#### a) Récolte de l'information

- Construction des requêtes.

Le point le plus important inhérent à la structure – ou la non-structure.. – d'Internet était *la pertinence des requêtes*. Ces requêtes devaient être définies de façon à ce que les résultats de recherche contiennent le moins de bruit possible.

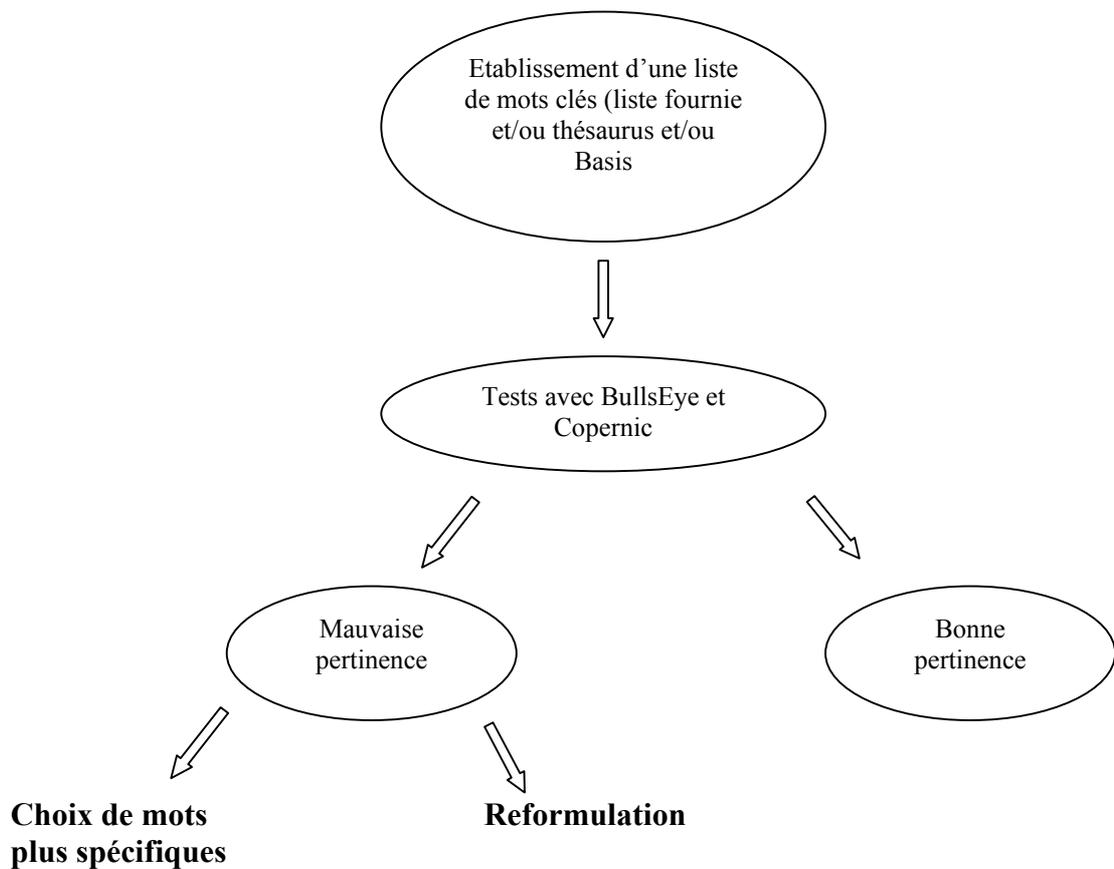
La première étape nécessaire a été la définition d'une liste de mots-clés. Pour ce faire, une collaboration avec le demandeur a été nécessaire, pour qu'il définisse avec la plus grande précision son sujet. Pour un laboratoire en particulier, ce fut le demandeur lui-même qui a fourni la liste des mots clés (Conditionnement), mais en règle générale, les mots clés étaient fournis :

- à partir de la demande,

- à partir d'un thésaurus (type CAB),
- à partir de la consultation directe de la base de données interne sous Basis.

Les requêtes élaborées à partir des mots clés ont dans un premier temps été testées avec Copernic 2000 et BullsEye, en parallèle. Malgré son dysfonctionnement, ce dernier pouvait s'avérer intéressant au niveau des résultats, et constituer un complément par rapport à Copernic.

La construction des requêtes a suivi le schéma suivant :



Voici le détail des tests pour chaque laboratoire :

*Nutrition*

*Polyphénols de fruits*

Etablissement d'une liste de mots clés à partir du sujet, consultation de Basis.



Tests sur BullsEye 2 et Copernic 2000



Bruit, manque de pertinence



Reformulation



**Tests sur Copernic : pertinence accrue**

*Conditionnement*

*“ Polymères barrières ”*

**Liste de mots clés fournie par le demandeur**



Tests sur BullsEye 2 et Copernic



Pertinence des résultats rapportés

*Chimie 5*

- *Mycotoxines : limites de tolérance dans les aliments*

**Mots clés déterminés à partir du sujet**



Tests sur Copernic uniquement



Pertinence, ajout de mots clés



- *Métaux lourds dans les jus de fruits et les boissons*



Tests sur Copernic uniquement



Choix de mots plus *spécifiques*

*Technologie*

*“ Alicyclobacillus acidoterrestris ” dans les jus de fruits*

Mots clés fournis par le demandeur



Tests sur Copernic et BullsEye



Pertinence des résultats

Des tests sont en cours avec *Glycérol et Acide gluconique dans le jus de raisin*. (Chimie 5).

- Taille des requêtes

Copernic est le logiciel le plus à même de répondre à la demande du CRPR, à l'heure actuelle, en termes d'automatisation des tâches.

Un bémol peut toutefois être ajouté en ce qui concerne la taille des requêtes : il est très vite apparu que Copernic était plus efficace avec des requêtes de 2 ou 3 mots au maximum.

Voici le détail d'une recherche et ses résultats en fonction du nombre de termes dans la requête – les chiffres désignent des mots :

|                                 |   |
|---------------------------------|---|
| “ 1 2 ” PRES (3 OU 4 OU 5 OU 6) | beaucoup de pages non pertinentes* (pages où ne figurent <i>que</i> 3 ou 4) |
| “ 1 2 ” PRES (3 OU 4)           | encore beaucoup de résultats incorrects                                     |
| “ 1 2 ” PRES 3                  | résultats pertinents  |

Ceci implique par voie de conséquence un nombre plus élevé de requêtes...

Au jour d'aujourd'hui, 9 “ profils-Internet ” ont été posés :

- Nutrition : 4
- Technologie : 1
- Conditionnement : 1
- Chimie : 3

Ils correspondent en tout à 22 requêtes définies. 5 ont été dernièrement proposées, et un laboratoire n'a pas encore donné de sujet (Microbiologie).

Sous Copernic, les requêtes sont classées par thèmes, sous la forme de calendriers de veille<sup>3</sup>.

- Chaque calendrier est paramétrable indépendamment des autres : heure, jour, mois, année.
- Les paramètres peuvent être à tout moment modifiés : cette flexibilité peut ainsi permettre ainsi de lancer des requêtes la nuit – ce qui permet de libérer le poste dans la journée , la connexion et la déconnexion au réseau étant parfaitement gérées.
- Les calendriers peuvent être inactivés (les recherches correspondantes sont suspendues)

a) Le traitement de l'information.

---

\* Ici, la pertinence est bien sûr entendue comme la pertinence de l'analyse effectuée par l'agent, non comme la pertinence par rapport à la demande.

<sup>1</sup> Voir copie d'écran en annexe

En théorie, Copernic analyse les résultats : il élimine les doublons, affiche un score de pertinence pour chaque document trouvé et effectue un tri selon ce score – par ailleurs, il surligne les mots clés trouvés et peut raffiner les résultats (c'est à ce stade que sont possibles l'utilisation du PRES et du SAUF).

La lecture fine des résultats de recherche peut cependant faire apparaître des erreurs d'analyse : certains doublons ne sont pas traités, des scores de pertinence bas peuvent être attribués à des documents très " pertinents ".

De fait, la pertinence de Copernic s'arrête où commence l'intelligence de l'humain, apte à déterminer par un survol de la page si l'information est *réellement* pertinente ou ne l'est pas.

Il ne sera vraisemblablement pas possible de relire tous les résultats un à un, pour vérifier si Copernic n'a pas fait d'erreur...

Par ailleurs, des analyses effectuées sous les yeux des demandeurs m'ont convaincu qu'il était plus de leur ressort que du mien de jeter une page à la corbeille : en effet, une recherche peut toujours rapporter des informations connexes, intéressantes pour le demandeur, et ne cadrant pas totalement avec la requête initialement convenue.

Concernant le traitement " humain " à venir, la solution consistera vraisemblablement en une relecture rapide des adresses de site - afin d'éliminer des doublons oubliés par Copernic – ainsi que du résumé édité pour chacun des résultats – afin d'éviter tous les sites " bruyants ".

Par ailleurs, afin d'accroître la pertinence des résultats, un affinage de la formulation sera également à prévoir, et ce en fonction des résultats obtenus dans le temps.

#### b) La diffusion de l'information.

Deux points sont à considérer :

D'une part, la politique du service de documentation est de fournir le moins de documents papier possible, afin de faciliter la gestion, et gagner de la place.

D'autre part, Copernic permet d'envoyer des rapports par e-mail. Ces rapports sont constitués de la liste des réponses obtenues, d'un titre, d'un cours résumé, et du score de pertinence pour chaque réponse. Ils comportent aussi et surtout les adresses cliquables de la page – adresse absolue et relative.

L'envoi des rapports par e-mail s'est avérée constituer la solution la plus simple pour fournir leurs résultats aux différents demandeurs, et ce pour trois raisons :

- l'absence d'un réseau constitué, mais la présence de postes Internet individuels,
- la lisibilité et l'ergonomie du rapport, permettant l'accès aux pages,
- l'autonomie des demandeurs face aux résultats diffusés : le demandeur imprime ce qui l'intéresse...mais garde la possibilité de consulter les résultats sauvegardés sur disque dur.

Pour l'un des demandeurs qui ne dispose pas de poste Internet, il est prévu de fournir les résultats sur disquette.

#### d) Le stockage de l'information

Rien n'est plus éphémère qu'une page Internet...

Les adresses de sites se modifient, changent d'adresse, ou disparaissent purement et simplement ! A nous les trop célèbres ERROR 404 ...

Devant ce constat de labilité d'Internet, il est nécessaire de prévoir un stockage de l'information.

Deux questions se sont posées :

1°) La place disponible à prévoir pour le stockage des pages :  
le résultat d'une recherche en moyenne multiplié par le nombre de requêtes et le nombre de mises à jour par an peut-il tenir sur un simple disque dur ?

2°) L'accessibilité :

Chacun devrait pouvoir être à même de consulter le résultat de sa recherche en plus du rapport envoyé sur le réseau. Internet n'étant pas en réseau, il fallait trouver un moyen pour que chacun accède aux pages correspondantes.

L'idée a donc été de transférer les pages sur le serveur, au côté des profils bibliographiques classiques, en plus de l'envoi de l'e-mail, qui n'offre que des liens.

Une classification par thèmes de ces profils est en cours, qui fait en particulier *intervenir la date* des mises à jour. Pour des profils classiques, les nouvelles références peuvent être copiées à la suite dans le même fichier Word. Pour des pages Internet, du fait de leur format, il est nécessaire de constituer des dossiers faisant apparaître la date et l'intitulé de la requête.

\*\*\*

## **V. Veille sur les Agents Intelligents.**

Le travail de Sandrine Royer, qui avait consisté à faire un état des lieux des agents disponibles s'était terminé en septembre 1999, et je suis arrivée exactement 7 mois après.

Dans l'intervalle, de nouveaux agents ont vu le jour, et pour les agents les plus sophistiqués, de nouvelles versions sont apparues.

Il s'est avéré nécessaire d'effectuer une veille technologique sur les agents.

Celle-ci a consisté à rester informé de l'actualité de ce domaine, par tous les moyens d'information mis à disposition dans le service :

- la lecture de la presse spécialisée : le service de documentation est, entre autres revues, abonné à *Netsources-Bases*, *Archimag*, *PC-Expert*, *Décision Micro*,<sup>4</sup>...
- la lecture d'ouvrages généraux, en particulier *Veille Stratégique sur Internet* de Carlo Revelli – édition 2000,
- l'abonnement à 3 listes de diffusion :
  - agents@egroups.fr
  - prospective@egroups.fr
  - moteurs@egroups.fren plus du suivi de celle de l'A.D.B.S.,
- la consultation de sites Internet spécialisés dans le domaine<sup>5</sup>,
- la visite du salon IDT les 23, 24 et 25 mai,
- la participation au séminaire de l'ADEST sur les " L'état de l'art dans le domaine des moteurs de recherche " et " l'évolution des problématiques de filtrage et des outils cartographiques : présentation de Neuroweb ", le 12 mai à l'Ecole des Mines de Paris.

Cette veille a revêtu un caractère formel, de même qu'un caractère informel, parfois:

- la lecture de livres et d'articles spécialisés, l'inscription à des listes de diffusion officielles a constitué la veille technologique formelle,
- les discussions avec d'autres utilisateurs (au sein de la promotion et ailleurs) ont constitué une veille technologique informelle, qui tout autant que les données diffusées par les médias ou les sociétés éditrices, ont permis de récolter des informations importantes.

Elle m'a permis de poursuivre la construction du répertoire des agents intelligents, en répertoriant 14 nouveaux agents :

- 5 aspirateurs de sites,
- 5 méta-moteurs et de super méta-moteurs de recherche,
- 4 outils d'analyses et de cartographies.

---

<sup>4</sup> Voir Bibliographie en annexe

<sup>5</sup> id.

Elle a aussi et surtout permis de ne “ passer à côté ” de l’agent idéal , qui serait apparu récemment...

## CONCLUSION

Automatiser les recherches pouvait relever de la gageure tant Internet est vaste, aléatoire, et tant il peut se révéler “ bruyant ” parfois.

Les agents intelligents offrent des possibilités très intéressantes dans l'optique de "dompter" le Web, mais, loin d'être parfaits, des solutions sont à apporter en terme de fiabilité.

Néanmoins, 4 laboratoires sur 5 qui en ont fait la demande ont des requêtes qui “ tournent ”, et les pages trouvées seront à terme disponibles sur le serveur.

Cette automatisation n'est pas terminée : des demandes ont été posées, mais les requêtes devront être modifiées au cours du temps en fonction des résultats qu'elles apporteront.

Par ailleurs, des "agents intelligents" apparaissent sans cesse; des solutions alliant la pertinence, la fiabilité, la rapidité et l'ergonomie apparaîtront peut-être, de même que des agents multi-services, alliant des fonctions de recherche et de surveillance.

Ce stage a été enrichissant du point de vue de la multiplicité des demandes émises, et du même coup, de la multiplicité des réponses à apporter.

Utiliser une grande variété d'outils (logiciels et supports informatiques) a également été très instructif. Un seul regret : celui de ne pas avoir eu le temps d'exploiter les logiciels de cartographie.

## ANNEXES

- 1. Bibliographie**
- 2. Glossaire**
- 3. Agents Intelligents**
- 4. Automatisation**
- 5. La gamme Pernod Ricard**

## ANNEXES

### **1. Bibliographie**

## Ouvrages

ANDRIEU, Olivier. *l'Officiel d'Internet*. Paris : Editions Eyrolles, 1995. 253 p. ISBN 2-212-08904

ANDRIEU, Olivier. *Méthodes et outils de recherche sur l'Internet*. Paris : Editions Eyrolles, 1997. 235 p. ISBN 2-212-08939-2

BONNET, Christophe et MACARY, Jean-François. *Technologies Push*. Paris : Editions Eyrolles, 1998. 253 p. ISBN 2-212-08991-0

*Guide pratique de la Veille Technologique sur Internet / Innovation 128*. Agence pour la Diffusion de l'Information Technologique, janvier 1998. 135 p.

KROL, Ed. *Le Monde Internet : guide & ressources*. 2<sup>e</sup> éd. Paris : Editions O'Reilly International Thomson, 1995. 543 p. ISBN 2-84177-000-1

*L'Internet professionnel : témoignages, expériences, conseils pratiques de la communauté, enseignement & recherche / Alain SIMERAY, Jean-Luc ARCHIMBAUD... et al..* Paris : CNRS Editions, 1995. 448 p. (CNRS & Universités). ISBN 2-271-05256

LARDY, Jean-Pierre. *Recherche d'information dans Internet : outils et méthodes*. Paris : ADBS Editions, 1996. 96 p. ISBN 2-901046-79-7

LELOUP, Catherine. *Moteurs d'indexation et de recherche : environnements client-serveur, Internet et Intranet*. Paris : Editions Eyrolles, 1998. 285 p. ISBN 2-212-08976-7

REVELLI, Carlo. *Intelligence stratégique sur Internet : comment développer des activités de veille et d'intelligence économique sur le web*. 2<sup>e</sup> éd. Paris : Dunod, 2000. 220 p. ISBN 2-10-005154-7

SAMIER, Henry et SANDOVAL, Victor. *La Recherche intelligente sur l'Internet : outils et méthodes*. Paris : Editions Hermès, 1998. 155 p. ISBN 2-86601-687-4

## Articles

BOUMENDIL, Pierre-Emmanuel. Google invente l'easy searching. *NetSources*, mars/avril 2000, N° 25, p. 13-15

CHANIAL, Emmanuel. Choisir le bon outil pour améliorer la veille sur Internet. *Décisions Micro&Réseaux*, 21 février 2000, N° 413, p. 54

DEROUET, Thierry. Protégez-vous d'Internet. *PC Expert*, août 2000, p. 105-117

FOENIX-RIOU, Béatrice. Strategic Finder, Copernic et les autres... *NetSources*, mars/avril 2000, N° 25, p. 1-11

FOENIX-RIOU, Béatrice. Web invisible : 550 milliards de documents ? *NetSources*, juillet/août 2000, N° 27, p. 1-3

GIMENEZ, Jean-Marc. L'aspirateur de sites, précieux pour la veille en ligne. *01 Informatique*, 19 mai 2000, N° 1589, p. 24

Internet, traquer les informations. *L'Ordinateur individuel*, décembre 1998, N° 101, p. 248-249

JOUINEAU, Marc. Internet, des agents pas si intelligents. *Archimag*, novembre 1998, N° 119, p. 30-32

ROUMIEUX, Olivier. Le discours de la méthode. *Archimag*, juin 2000, N° 135, p. 26-28

ROUMIEUX, Olivier. Les tendances du prêt-à-chercher. *Archimag*, juin 2000, N° 135, p. 24-25

TRAN, Pierre. Trouver le meilleur du Net. *PC Expert*, juin 2000, p. 167-178

#### Sites et pages Internet

ABCDOC (France). (page consultée le 17 août 2000). *Lettre hebdomadaire d'info sur les outils du Net*. [en ligne]

<http://www.abcdoc.net/hebdoc/>

AGENTLAND (France) (page consultée le 17 août 2000) [site en construction]

<http://www.agentland.com>

FOUNDATION FOR INTELLIGENT PHYSICAL AGENTS (Suisse). (page consultée le 17 août 2000) [en ligne]

<http://www.fipa.org/inform/issue1.pdf>

IDF.NET. (page consultée le 17 août 2000). *Liste complète des aspirateurs de site*. [en ligne]

<http://www.idf.net/info/table.html>

MIT MEDIA LAB (Etats-Unis). (page consultée le 17 août 2000) *Software agents group*. [en ligne]

<http://agents.www.media.mit.edu/groups/agents/>

SWEDISH INSTITUTE OF COMPUTER SCIENCES (Suède). (page consultée le 17 août 2000) *Agent based systems*. [en ligne]

<http://www.sics.se/isl/abc/survey.html>

THE AGENT SOCIETY ORGANIZATION. (page consultée le 17 août 2000) *Agent product and research activities*. [en ligne]

<http://www.agent.org/>

THE UNIVERSITY OF MARYLAND – BALTIMORE COUNTY (Etats-Unis). *UMBC Agent Web*. [en ligne]

<http://agents.umbc.edu/>

VEILLE.COM (France). (page consultée le 17 août 2000) *Le laboratoire des agents intelligents* [en ligne]  
<http://www.veille.com/>

## ANNEXES

### **2. Glossaire**

|                       |   |
|-----------------------|---|
| <b>Agent d'alerte</b> | Appelés aussi <i>e-mail notifiers</i> ou <i>notification agents</i> , ils envoient un courrier électronique chaque fois qu'une page Web sélectionnée est modifiée. Il est pour certains agents d'alerte possible de filtrer les changements survenus en utilisant des mots clés appropriés.   |
| <b>Aspirateur</b>     | Agent permettant la copie sur disque dur de sites ou de pages Web. Les mises à jour peuvent être programmées, et couplées à des alertes quand le contenu d'une page est modifié.  |
| <b>Métamoteur</b>     | Le principe des métamoteurs est d'aller plus loin ( <i>meta</i> ) des moteurs traditionnels. Il s'agit d'un hybride pouvant interroger plusieurs outils simultanément (répertoires et moteurs).   |
| <b>Moteur</b>         | C'est un outil puissant permettant la recherche précise d'informations sur Internet. Il est constitué de deux éléments : un robot (ou <i>spider</i> ), qui visite en permanence des millions de documents, afin de les stocker ; un système d'indexation (propre à chaque robot) permettant l'analyse de l'information afin que les utilisateurs puissent la trouver en utilisant les mots-clés appropriés. |
| <b>Plugg-in</b>       | Un fichier plugg-in contient de nouvelles catégories de bases de données interrogeables, qui peuvent rassembler des sources d'information couvrant un secteur ou un thème commun (biotechnologies, presse économique, agronomie...)   |
| <b>Pull</b>           | De l'anglais <i>pull</i> signifiant "tirer". Démarche active consistant à tirer l'information à soi : ex : effectuer une recherche via un moteur, déléguer une recherche à un agent de recherche,   |
| <b>Push</b>           | De l'anglais <i>push</i> signifiant "pousser". L'information est poussée vers nous, laissant l'utilisateur passif. On parle aussi de <i>push media</i> , <i>webcasting</i> , <i>webdiffusion</i> , <i>broadcasting</i> , etc..  |
| <b>Spider</b>         | Automate inclus dans un moteur de recherche, dont le rôle est de visiter les pages Web, de les lire, et de suivre les liens vers d'autres sites.  |
| <b>Web visible</b>    | Ensemble des pages Web accessibles par les moteurs de recherche traditionnels. <i>Voir aussi Web invisible.</i>   |
| <b>Web invisible</b>  | Base de données indépendantes, très riches, et qui ne peuvent être atteintes par les robots de recherche généralistes.<br>NB : leur éparpillement rend leur utilisation souvent longue et difficile.  |

## ANNEXES

### **3. Agents Intelligents**

NOM

**BULLS EYE 2.5 Pro**

|                         |   |
|-------------------------|---|
| <b>DÉVELOPPEUR</b>      | Intelliseek   |
| <b>PARTICULARITÉS</b>   | Agent de collecte d'informations et de surveillance de site   |
| <b>FONCTIONS</b>        | Recherche:+++<br>Indexation:+++<br>Filtrage :++<br>Présentation:++<br>Distribution:++<br>Aide à la décision   |
| <b>CARACTÉRISTIQUES</b> | Les mêmes que la précédente, avec en plus : <ul style="list-style-type: none"><li>- 800 sources du Web visible et invisible</li><li>- automatisation des recherches selon un calendrier paramétrable : jour, mois, année</li><li>- possibilité de surveiller une page Web, avec un système d'alerte (e-mail, téléphone portable..)</li><li>- génération de rapports</li><li>- une interface améliorée</li></ul> |
| <b>SORTIE</b>           | disponible  |
| <b>CONFIGURATION</b>    | Windows 95/98/NT<br>N'est pas compatible avec la version 5.5 d'Internet Explorer  |
| <b>MAINTENANCE</b>      | Support@intelliseek.com.  |
| <b>POINTS POSITIFS</b>  | Puissance des recherches et rapidité  |
| <b>POINTS NÉGATIFS</b>  | - des bugs dans le logiciel empêchent le déclenchement des recherches automatiques, ou la surveillance des pages (Tracker) - constaté chez d'autres utilisateurs également.<br>- La mise à jour des recherches n'est pas prévue lors de l'installation<br>- service technique difficile à joindre par e-mail  |
| <b>PRIX</b>             | 249 \$  |
| <b>CONTACT</b>          | support@intelliseek.com   |
| <b>DEJA UTILISE PAR</b> |   |
| <b>TÉLÉCHARGEMENT</b>   | À partir du site  |
| <b>DISTRIBUTEURS FR</b> | non   |
| <b>DÉMONSTRATIONS</b>   |   |
| <b>SALONS</b>           |   |

0 : fonction non remplie  
+ : fonction optionnelle ou peu remplie

++ : fonction remplie  
+++ : fonction principale

**DÉVELOPPEUR  
PARTICULARITÉS**

Agent Technologies Inc.  
Agent de collecte et agent de veille.  
Méta-moteur off-line

**FONCTIONS**

Recherche:+++  
Indexation:++  
Filtrage :+++  
Présentation:  
Distribution:  
Aide à la décision:

**CARACTÉRISTIQUES**

Les mêmes que dans la version 99, avec en plus ::

- Vérificateur d'orthographe
- Traduction possible des documents trouvés (14 paires de langues)
- Opérateur logique PRES ajouté dans la fonction "Raffiner"
- Assistant de création d'un Calendrier de Veille, permettant l'automatisation de recherches ciblées, selon la périodicité voulue.
- Automatisation de l'envoi des rapports par e-mail, à la périodicité voulue.
- Regroupement possible de plusieurs recherches dans un même calendrier de veille
- Ordonnancement possible des recherches à l'intérieur des recherches
- Ouverture et fermeture automatique de la connexion à Internet par modem, au début et à la fin d'un calendrier de veille.
- Tâche automatisée permettant le téléchargement des documents.
- Plusieurs moteurs de recherche ajoutés dans les domaines de recherche existants, et deux domaines de recherche standard ajoutés : en tout 610 moteurs accessibles.
- 7 nouveaux domaines de recherche optionnels, liés à la langue ou au pays.
- Personnalisation de l'interface par possibilité de changer les skins (habillages) téléchargeables
- Amélioration de la boîte de dialogue

**SORTIE  
CONFIGURATION**

disponible

- Windows 95/98/NT
- S'intègre à Internet Explorer

**MAINTENANCE**

- <http://www.copernic.com/fr/support>

- POINTS POSITIFS** - convivialité alliée à l'efficacité
- POINTS NÉGATIFS** - taille des requêtes limitée, possibilité d'utiliser l'opérateur PRES à la phase de raffinage uniquement.

**PRIX** 79,95 \$ US (env.480 F)

**CONTACT**

**DEJA UTILISE PAR** De nombreux utilisateurs...  
**TÉLÉCHARGEMENT** A partir du site.  
**DISTRIBUTEURS FR**  
**DÉMONSTRATIONS**  
**SALONS**

0 : fonction non remplie  
+ : fonction optionnelle ou peu remplie

++ : fonction remplie  
+++ : fonction principale

NOM

**STRATEGIC FINDER 1.5**

**DÉVELOPPEUR**  
**PARTICULARITÉS**

Digimind  
Agent de recherche d'information sur le Web, destiné à la veille stratégique.  
Mise à jour de la version 1.3

**FONCTIONS**

Recherche:+++  
Indexation:++  
Filtrage :++  
Présentation:0  
Distribution:0  
Aide à la décision:0

**CARACTÉRISTIQUES**

Les mêmes que la version 1.3, avec en plus :

- Un nombre accru de sources d'information (1000 dans la version complète)
- Une interface plus conviviale
- L'édition de nouveaux plugg-in :
- Biognome : plus de 500 bases dans le domaine des biotechnologies.
- AgroFinder : plus de 110 bases de données dans le domaine de l'industrie agro-alimentaire, dans les secteurs suivants: institutions, presse, sites officiels, sécurité et qualité, sites scientifiques, sites scientifiques, associations, législation,...

Disponible en français et en anglais

**SORTIE**

Depuis le 14 janvier 2000

**CONFIGURATION**

Windows 95/98/NT

**MAINTENANCE**

Support@digimind.fr

**POINTS POSITIFS**  
**POINTS NÉGATIFS**

- Le nombre de sources, l'installation des plugg-in
- perte du NEAR
- pas d'automatisation des recherches possibles alors que les recherches sont plutôt longues
- 

**PRIX**

457 euros (2970 F)

**CONTACT**

Patrice François  
Digimind  
12, rue Ampère  
BP 267  
38 016 Grenoble Cedex 01  
Tél : +33 (0)4 76 70 13 78  
Fax : +33 (0)4 76 70 96 53

patrice.francois@digimind.fr  
support@digimind.fr

**DEJA UTILISE PAR**

Thomson CSF, Total Fina Elf, France Telecom, Le Point, CNRS, Danone, Rhodia, ...

**TÉLÉCHARGEMENT  
PARTENAIRES**

A partir du site

ATLANTIC INTELLIGENCE  
<http://www.alertia.com>

COMPETITIVE INTELLIGENCE MANAGEMENT  
<http://www.cim.-intelligence.com>

CYBIL

INFORIZON

KEYWORDS - Diary Ba Van Tilbeurgh

SENEQUE  
<http://www.seneque.fr>

STRATEGIC ROAD - Your competitive intelligence  
directory  
<http://www.strategic-road.fr>

VEILLE MAGAZINE  
<http://veillemag.com>

**DÉMONSTRATIONS**

Atlantic Intelligence (pour AgroFinder )  
1, rue Louis Marin  
44 200 Nantes  
Tél : 02 51 72 98 18  
Fax : 02 51 72 98 15

**SALONS**

Salon IDT 2000

0 : fonction non remplie  
+ : fonction optionnelle ou peu remplie

++ : fonction remplie  
+++ : fonction principale

## I. Liste des moteurs interrogés par Copernic 2000 et BullsEye sur le Web

| Copernic 2000   | BullsEye 2 Pro   |
|---|--|
| <p style="text-align: center;"> AltaVista<br/> AOL.COM search<br/> Direct Hit<br/> EuroSeek<br/> Excite<br/> FAST Search<br/> Google<br/> GoTo.com<br/> HotBot<br/> Infoseek<br/> LookSmart<br/> Lycos<br/> Magellan<br/> MSN Web Search<br/> Nescape Netcenter<br/> Open Directory Project<br/> Snap<br/> WebCrawler<br/> Yahoo </p> | <p style="text-align: center;"> About.com<br/> All The Web<br/> AltaVista Advanced<br/> AltaVista Advanced Canada<br/> AOL NetFind Kids Only<br/> Britannica's Internet Guide<br/> Direct Hit Web Search<br/> Disney's Internet Guide Search<br/> Dow Jones Business Directory<br/> EuroFerret<br/> EuroSeek<br/> Excite<br/> Excite Selected<br/> Google<br/> Goto.com<br/> HotBot<br/> InfoHiway<br/> InfoSeek<br/> LookSmart<br/> Lycos<br/> Lycos Personal Homepage<br/> Magellan<br/> MSN Web Search<br/> NerdWorld Internet Subject Index<br/> Search<br/> Open Directory<br/> Open Text<br/> Scrub The Web<br/> Snap<br/> The Internet Resource Exchange<br/> Thunderstone Web Search<br/> TradeWave Galaxy<br/> USA Today Hotsites<br/> WebCrawler<br/> Yahoo! Advanced<br/> Yahooligans! </p> |

## II. Les autres sources de Copernic 2000 Pro

|   |  |
|---|--|
| <p><b>Le Web français</b></p> <p>Altavista Français<br/>Breizoo<br/>Ecila<br/>EurSeek<br/>Excite<br/>FAST Search<br/>FRANCITE.COM<br/>Google<br/>HotBot<br/>InfoSeek France<br/>La Toile du Québec<br/>Lokace</p> <p>Lycos France<br/>Nomade<br/>Snap<br/>Spray.fr<br/>Sympatico-Lycos<br/>Voilà<br/>Yahoo!France</p> | <p><b>Le Web Portugal</b></p> <p>AIEOU<br/>AlgarVio<br/>BemVindo a Voilà<br/>Direct Hit<br/>EuroSeek Portugal<br/>FAST Search<br/>HotBot Portugal<br/>Lycos Portugal<br/>NetINDEX<br/>Portal Busca<br/>Portugal INFO<br/>SAPO - Portugal<br/>Online<br/>Snap</p> |
|---|--|

|  |  |
|--|--|
| <p><b>Le Web Australie</b></p> <p>Altavista<br/>Anzak Search<br/>Anzwers<br/>Cowley Internet<br/>Direct Hit<br/>Excite<br/>FAST SEARCH<br/>HotBot<br/>Infoseek<br/>LookSmart<br/>OzSearch<br/>Snap</p> | <p><b>Le Web - UK</b></p> <p>Direct Hit<br/>EuroSeek<br/>Excite UK<br/>FAST Search<br/>HotBot<br/>InfoSeek<br/>LineOne<br/>LookSmart<br/>Lycos UK<br/>Mirago<br/>SearchEngine.com<br/>Snap<br/>Snoopa<br/>UK Directory</p> |
|--|--|

|   |  |
|---|--|
| <p>WebSearch AU<br/>Yahoo! Australia &amp; NZ</p> | <p>UK Plus<br/>WebSearch UK<br/>Yahoo! UK&amp;Ireland<br/>Yell</p> |
|---|--|

|   |  |
|---|--|
| <p><b>Le Web Canada</b></p> <p>Altavista Canada<br/>AOL.CA NetFind<br/>Canada.com<br/>Direct Hit<br/>Excite Canada<br/>Fast Search<br/>HotBot<br/>Infoseek<br/>LookSmart<br/>Lycos<br/>Snap<br/>Yahoo! Canada</p> | <p><b>Le Web Allemand</b></p> <p>Abacho.com<br/>Altavista<br/>Crawler.de<br/>DINO<br/>Direct Hit<br/>EuroSeek<br/>Excite<br/>FAST Search<br/>Fireball<br/>Google<br/>HotBot<br/>Infoseek<br/>Lycos<br/>Snap<br/>Spray.net<br/>Yahoo!</p> |
|---|--|

|   |   |
|---|---|
| <p><b>Le Web Espagnol</b></p> <p>Altavista Magellanes<br/>Altavista Spanish<br/>Antena<br/>Buscador.clarin<br/>Elcano<br/>EuroSeek Spanish<br/>Excite Spain<br/>FAST Search<br/>Google<br/>HotBot<br/>InfoSeek Spain<br/>Mexmaster<br/>Olé<br/>Snap<br/>Yahoo!<br/>Yupi</p> | <p><b>Le Web Italien</b></p> <p>Altavista<br/>Arianna<br/>Direct Hit<br/>EuroSeek<br/>Excite<br/>Fast Search<br/>Google<br/>HotBot<br/>Infoseek<br/>Katalogo<br/>La Republica<br/>Lycos<br/>Ragno Italiano<br/>Snap<br/>Spray.it<br/>Virgilio</p> |
|---|---|

|  |        |
|--|--------|
|  | Yahoo! |
|--|--------|

|  |   |
|--|---|
| <p style="text-align: center;"><b>Sciences</b></p> <p style="text-align: center;">Astronomy magazine<br/>Discovery Channel Online<br/>EEVL<br/>Exn<br/>NASA<br/>National Geographic.com<br/>New Scientist<br/>Newswise<br/>ScienceDaily Magazine<br/>Scientific American</p> | <p style="text-align: center;"><b>Santé</b></p> <p style="text-align: center;">aHealthyMe<br/>Ask Dr. Weil<br/>Drkoop.com<br/>DrugInfoNet<br/>HealthAnswers<br/>HealthAtoZ<br/>HealthFinder<br/>InfoseekGoHealth<br/>Mayo Clinic Health Oasis<br/>MedExplorer<br/>MedicineNet.com<br/>MediConsult.com<br/>MEDLINEplus<br/>OnHealth<br/>The Thrive Health Library<br/>WebMD<br/>YourHealth.com</p> |
|--|---|

|   |   |
|---|---|
| <p style="text-align: center;"><b>Groupes de discussion</b></p> | <p style="text-align: center;">AltaVista - Discussion Group Search<br/>CENT Help.com<br/>Deja.com</p> |
|---|---|

**Autres domaines :**

**Adresses de courrier**

**Achat-livres**

**Achat - logiciels**

**Achat - Matériel Informatique**

**Affaires et finances**

**Articles sur les techno:logies**

**Automobiles**

**Cinéma**

**Enchères**  
**Encyclopédies**  
**Enfants**  
**Fils de presse**

### III. Les autres sources de BullsEye 2000 Pro

#### News

|   |  |
|---|--|
| ABC News                                | Ivanhoe's Medical Breakthroughs Search         |
| Altavista Advanced Canada News Search   | LAN Times Search                               |
| Altavista Canada News Search            | Life Magazine Search                           |
| AnchorDesk Search                       | Los Angeles Times Web Site Search              |
| Associated Press 14-days Archive        | MacWEEK Search                                 |
| Baltimore Sun 14 Day Search             | MacWorld Search                                |
| BBC News OnLine Search                  | Miami Herald Recent Issue Search               |
| BBC Website Search                      | Minneapolis Star Tribune 21 Day Search         |
| Boston Globe Current Issue Search       | Money Magazine Search                          |
| BusinessWire Search                     | Open Systems Today Search                      |
| Byte Magazine Site & Magazine Archive   | PBS Online News HourSearch                     |
| Canada Newswire Search by Keyword       | PC Computing Search                            |
| CIO and WebMaster Magazine Search       | PC Magazine Search                             |
| CMP Technology Database Search          | PC Week Search                                 |
| CNN Interactive Search                  | PC Week Search                                 |
| CNNfn Searcher                          | PC WorldOnline Search                          |
| Communications Week Int. Search         | PlugIn dDatamation Search                      |
| Communications Week Search              | PR NewsWire Search                             |
| Computer Shopper Search                 | ReutersHealth HealthLine Search                |
| Data Communications Search              | ReutersHealth Professional Medical News Search |
| Detroit Free Press Current Issue search | Sacramento Bee 7 day Search                    |
| Electric Library News Search            | San Jose Mercury News                          |
| Electronic Buyers News Search           | Seattle Times 7 day search                     |
| Electronic Engineering Times Search     | TechInvestor Search                            |
| Entertainment Weekly Search             | TechWeb News Search                            |
| EurekaAlert! Search                     | The Medical Tribune Search                     |
| Excite News Tracker                     | The Medical Tribune Search                     |
| Family PC Search                        | The Red Herring Magazine Archive               |
| Fast Company Web Archives Search        | Time Daily Search                              |
| Forbes ASAP Search                      | Time Magazine Search                           |
| Forbes Digital Tool Search              | Upside Online Search                           |
| Fortune Magazine Search                 | URL Wire                                       |
| Gina Internet Wire Archive              | USA Today News search                          |
| IDG.net Global Search                   | US News Back Issues search                     |
| Inc. Magazine Online Search             | US News Online Search                          |
| Information Week Search                 | Washington Post 14 Day Search                  |
| InfoSeek Newsq Wires                    | WebReview Search                               |
| InfoWorld Electric Search               | Windows Magazine Search                        |
| InteliHealth.com News Search            | Worth Magazine Search                          |
| Inter@ctive Week Search                 | Yahoo! News Search                             |
| Interactive Age Search                  | Your Company Online Search                     |
| Internet Week Search                    | Your Health daily search                       |
| Internet World Search                   | ZDNN Search                                    |
| InternetNews.com Search                 | ZDTV Search                                    |



|   |  |
|---|--|
| <p style="text-align: center;"> Ask Dr. Weil Search<br/> Doctor's Guide Medical<br/> News and Conference<br/> Database Search<br/> Drug Infonet<br/> HealthAnswers.com Search<br/> InteliHealth.com Ask The<br/> Doc Search<br/> InteliHealth.com Search<br/> <br/> <b>Health</b><br/> Karolinska Institute<br/> Look It Up Database Search<br/> <br/> MAYO Health O@sis<br/> Search<br/> MedExplorer<br/> MediConsult.com Search<br/> National Library Of<br/> Medicine<br/> Parentsplace.com Search<br/> Parent Time Search<br/> Thrive@ Health Search </p> | <p style="text-align: center;"> <br/> DejaNews Internet Newsgroup<br/> Standard<br/> DejaNews Internet Newsgroups<br/> Search<br/> RemarQ Messages Search<br/> Silicon Investor Message Bords<br/> Search<br/> The Motley Fool Message Boards<br/> Search<br/> Yahoo! Message Boards Search </p> <p><b>Discussions</b></p> |
|---|--|

**Autres sources :**

- Shopping**
- Multimédia**
- Jobs**
- Books**
- Software**
- Computers**
- References**
- Entertainment**
- College**
- International**

NOM

**WEBSPECTOR**

|                         |   |
|-------------------------|---|
| <b>DÉVELOPPEUR</b>      | Illumix   |
| <b>PARTICULARITÉS</b>   | Aspirateur de pages de deuxième génération (alerte off-line)  |
| <b>FONCTIONS</b>        | Recherche: 0<br>Indexation: 0<br>Filtrage :+++<br>Présentation:++<br>Distribution:<br>Aide à la décision:   |
| <b>CARACTÉRISTIQUES</b> | <ul style="list-style-type: none"><li>- " Un des meilleurs agents disponibles pour mener des activités de veille" in. "Intelligence Stratégique sur Internet"- Carlo Revelli - Dunod Ed.</li><li>- Puissantes fonctions de filtrage et de compararison, qui permettent la mise en évidence de nouveautés apparaissant sur une page Web.</li></ul> |
| <b>SORTIE</b>           | Disponible.   |
| <b>CONFIGURATION</b>    | Windows 95/98/NT  |
| <b>MAINTENANCE</b>      | <a href="http://www.illumix.com/support.htm">www.illumix.com/support.htm</a>  |
| <b>POINTS POSITIFS</b>  | <ul style="list-style-type: none"><li>- Veille par mots clés et mise en valeur des changements directement sur la page html, de manière très visuelle.</li><li>- Les fichiers sont enregistrés au format d'origine et peuvent donc être éventuellement indexés ou traités avec d'autres logiciels.</li><li>- Simple et intuitif</li></ul>         |
| <b>POINTS NÉGATIFS</b>  | Impossibilité de surveiller le site en entier, il faut créer un agent pour chaque page à surveiller.  |
| <b>PRIX</b>             | Entre 50 et 300 \$ en fonction des versions.  |
| <b>CONTACT</b>          | <a href="http://www.illumix.com">www.illumix.com</a>  |
| <b>DEJA UTILISE PAR</b> |   |
| <b>TÉLÉCHARGEMENT</b>   | À partir du site  |
| <b>DISTRIBUTEURS FR</b> | non   |
| <b>DÉMONSTRATIONS</b>   |   |
| <b>SALONS</b>           |   |

0 : fonction non remplie  
+ : fonction optionnelle ou peu remplie

++ : fonction remplie  
+++ : fonction principale

## ANNEXES

### **4. Automatisation**

## ANNEXES

### **5. La gamme Pernod Ricard**