



HAL
open science

Choix d'un agent intelligent et mise en place d'une application push

Aurélie Gondouin

► **To cite this version:**

Aurélie Gondouin. Choix d'un agent intelligent et mise en place d'une application push. domain_shs.info.docu. 2000. mem_00000073

HAL Id: mem_00000073

https://memsic.ccsd.cnrs.fr/mem_00000073

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.



Hoechst Marion Roussel

Service de Documentation et d'Information Scientifiques du site de Romainville
102, route de Noisy
93235 Romainville Cedex



Choix d'un agent intelligent et mise en place d'une application push.

Rapport de stage réalisé par **Aurélie GONDOUIN**.

Sous la tutelle de **M. Jean-Marc GIRODEAU**.

15 mai – 30 septembre 2000.



Résumé

Parmi les sites de la société Aventis, issue de la fusion de Hoechst et de Rhône-Poulenc, celui de Romainville possède un service de documentation et d'information scientifiques performant. Pour développer l'utilisation d'Internet / Intranet, ce service a souhaité acquérir un agent intelligent, et mettre en place une application push. Mais qu'est vraiment un agent intelligent : un produit de l'intelligence artificielle ? la dernière génération d'outils de recherche ? Les agents commercialisés peuvent avoir quatre fonctionnalités différentes : recherche, aspiration, traitement, et veille. Après avoir recensé les besoins des utilisateurs, et les outils disponibles sur le marché, des évaluations ont conduit à faire ressortir trois agents susceptibles d'être acquis par le service.

Mots-clés

- agent intelligent
- push
- Internet
- web
- industrie pharmaceutique
- outil de recherche
- veille
- recherche d'informations
- aspiration
- évaluation

Remerciements

Je tiens à remercier particulièrement :

M. Jean-Marc GIRODEAU, pour son encadrement et ses conseils, et surtout pour m'avoir permis de conduire un projet dans son intégralité,

l'ensemble du personnel du service de Documentation et d'Information Scientifiques du site de Romainville, pour leur accueil chaleureux et leur présence,

Claire MAURICE, pour son soutien et sa bonne humeur,

plusieurs étudiants du DESS, pour leur collaboration.

<h2>SOMMAIRE</h2>

Résumé et mots-clés	2
Remerciements	3
SOMMAIRE	4
INTRODUCTION	6
I. Présentation de l'entreprise	7
1. La fusion Hoechst / Rhône-Poulenc	7
<i>a. Hoechst</i>	7
i. Organigramme de Hoechst au moment de la fusion.....	7
ii. Hoechst Marion Roussel.....	7
<i>b. Rhône-Poulenc</i>	8
i. Organigramme de Rhône-Poulenc au moment de la fusion.....	8
ii. Branche Pharma de Rhône-Poulenc.....	8
<i>c. La fusion</i>	8
2. Le service de Documentation et d'Information Scientifiques du site de Romainville (DIS)	9
<i>a. La bibliothèque scientifique</i>	9
<i>b. La gestion des données et des connaissances</i>	9
<i>c. La veille scientifique et concurrentielle</i>	10
3. Objectifs du stage au sein de l'entreprise	10
II. Agents intelligents : définition	11
1. En théorie : le point de vue scientifique	11
2. En pratique : les outils du marché	12
<i>a. Les précurseurs</i>	13
i. Les annuaires.....	13
ii. Les moteurs de recherche.....	13
iii. Les métamoteurs de recherche.....	13
<i>b. Agents intelligents</i>	14
i. Recherche d'informations.....	15

ii. Aspiration de sites.....	15
iii. Traitement des informations recueillies.....	16
iv. Veille.....	17
III. Recensement des besoins.....	18
1. Les besoins du service de Documentation et d'Information Scientifiques.....	18
<i>a. Recherche d'informations.....</i>	<i>18</i>
<i>b. Veille.....</i>	<i>19</i>
2. Les besoins des "end-users".....	19
IV. Evaluation.....	20
1. La grille d'évaluation.....	20
2. Les outils à tester.....	22
<i>a. Les contraintes.....</i>	<i>22</i>
<i>b. Les outils.....</i>	<i>23</i>
3. Les tests.....	26
V. Résultats.....	28
1. Problèmes rencontrés.....	28
2. Choix d'un agent intelligent.....	29
CONCLUSION.....	30
BIBLIOGRAPHIE.....	31
GLOSSAIRE.....	35
ANNEXES.....	37

NB : - [1] (par exemple) renvoie à la bibliographie p.31
- [*] renvoie au glossaire p.35

INTRODUCTION

Depuis sa naissance à la fin des années 60, Internet n'a cessé de croître, tant en nombre d'utilisateurs qu'en nombre de sites accessibles. Aujourd'hui, il est devenu l'une des principales sources d'informations. Hormis les milliards de pages consultables, il permet désormais d'accéder à des bases de données et même des périodiques.

A l'heure actuelle, Internet s'avère être une ressource incontournable pour une entreprise comme Aventis. Il regorge en effet d'informations intéressantes : scientifiques bien sûr (articles, congrès, ...), mais aussi concurrentielles (nouvelles concernant l'industrie pharmaceutique, renseignements sur des entreprises du domaine, ...).

Cependant, bien que les informations pertinentes fourmillent, il n'est pas toujours aisé de les retrouver. C'est pourquoi les acteurs d'Internet mettent à notre disposition une pléiade d'outils de recherche (annuaires et moteurs), couvrant chacun une partie du web. Ils peuvent retrouver un grand nombre de pages, mais pas la totalité.

Pour développer l'utilisation d'Internet, le service de documentation et d'information scientifiques du site de Romainville a souhaité acquérir un "agent intelligent". Un outil de ce type offre par exemple la possibilité d'interroger simultanément plusieurs outils de recherche, et donc d'augmenter le nombre de documents retrouvés (grâce à l'addition des parties du web couvertes par chaque outil). En outre, un agent intelligent peut proposer d'autres fonctionnalités ou options, comme la veille de sites.

Après une présentation de la société Aventis et du service de Documentation et d'Information Scientifiques du site de Romainville, je détaillerai les objectifs de mon stage.

Puis, je tenterai de définir ce qu'est un agent intelligent, du point de vue scientifique et commercial.

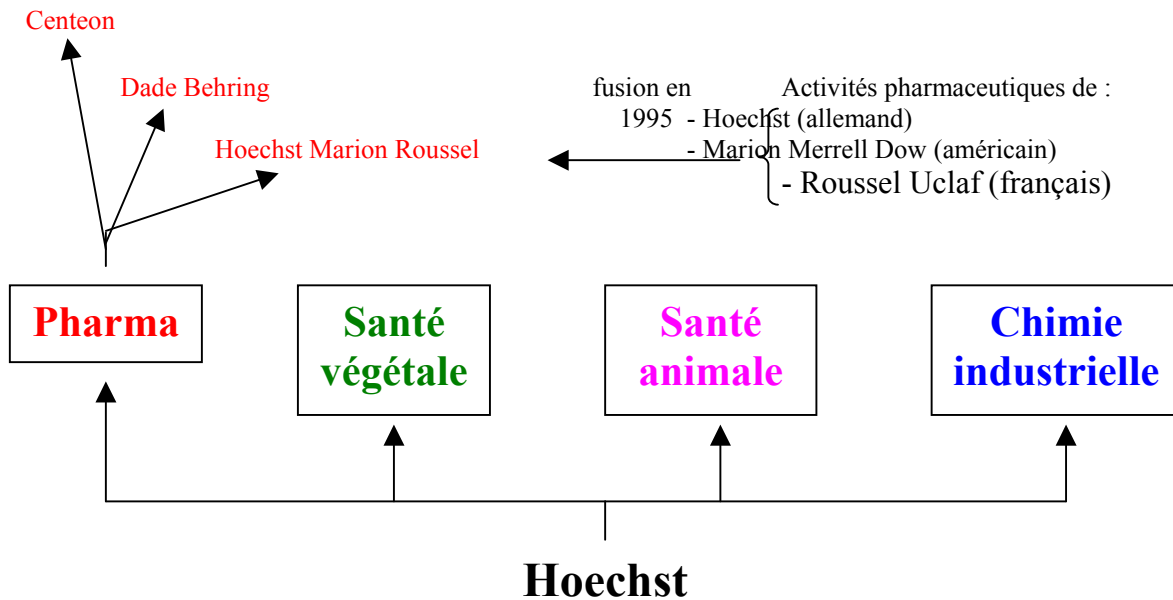
Enfin, j'exposerai l'ensemble de mon étude, du recensement des besoins du site jusqu'à mes résultats.

I. Présentation de l'entreprise

1. La fusion Hoechst / Rhône-Poulenc

a. Hoechst

i. Organigramme de Hoechst au moment de la fusion



ii. Hoechst Marion Roussel

Le site de Romainville concernait Hoechst Marion Roussel jusqu'à la fusion. Les activités d'aujourd'hui sont les mêmes qu'en 1998.

L'organisation s'appelle *Drug Innovation and Approval* (Innovation Thérapeutique et Enregistrements), soit DI&A. Quatre aspects sont pris en compte :

- identification des candidats cliniques
- développement de médicaments
- enregistrement des médicaments
- pharmacovigilance.

Au centre de recherche de Romainville, la DI&A s'appuie sur deux domaines, c'est-à-dire deux "Groupes de maladies" (*Disease Groups*) :

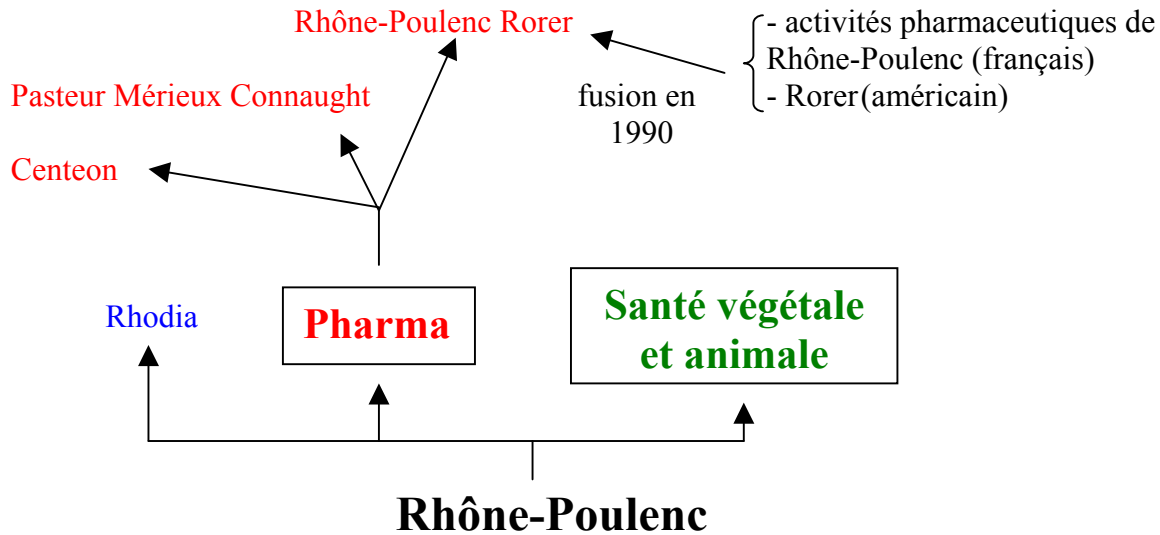
- Maladies infectieuses
 - antibiotiques : résolution des principaux problèmes de résistance développée par des souches bactériennes
 - antifongiques : attaque des infections opportunistes qui atteignent les patients immunodéprimés
- Maladies de l'os : prévention et guérison des maladies qui deviennent de plus en plus fréquentes du fait du vieillissement de la population.

Ces recherches font appel à différents centres d'expertise (chimie, biotechnologies, ...).

En 1998, Hoechst Marion Roussel se place au 6^{ème} rang européen des industries pharmaceutiques, et est un des premiers groupes pharmaceutiques étrangers sur les marchés japonais et latino-américains. En France, il est le premier producteur de médicaments.

b. Rhône-Poulenc

i. Organigramme de Rhône-Poulenc au moment de la fusion



ii. Branche Pharma de Rhône-Poulenc

Comme chez Hoechst, on retrouve Centeon. Cette société est en effet commune aux deux entreprises (50 % pour chacune). Centeon est classée au deuxième rang mondial en ce qui concerne les protéines plasmatiques.

Pasteur Mérieux Connaught est le leader mondial pour les vaccins humains, qui immunisent en particulier la méningite, la grippe, la coqueluche et l'hépatite A.

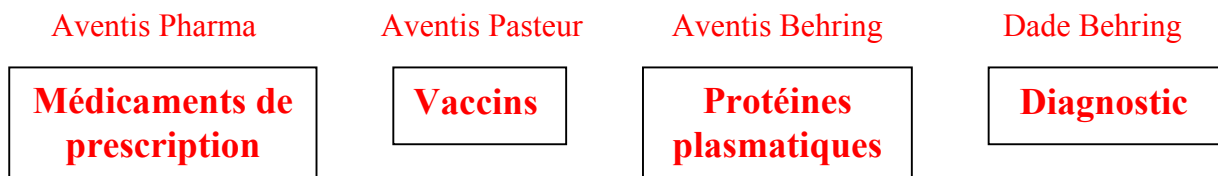
Rhône-Poulenc Rorer, quant à lui, est spécialisé dans les médicaments (découverte, développement, production, commercialisation). Il couvre trois domaines :

- Thrombose / Cardiologie
- Oncologie
- Respiratoire / Allergie.

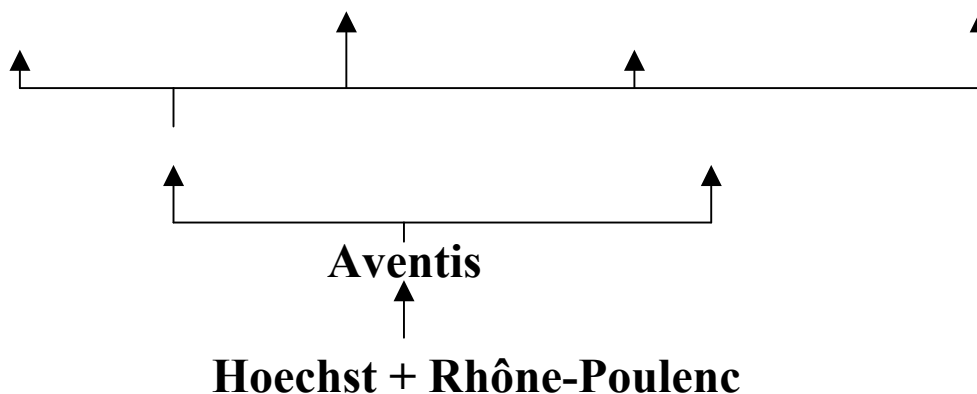
En 1998, il est le premier groupe pharmaceutique en France, et est placé au septième rang en Europe.

c. La fusion

Annoncée le 1^{er} décembre 1998, la fusion entre Hoechst et Rhône-Poulenc s'est confirmée le 17 mai 1999.



Choix d'un agent intelligent et mise en place d'une application push.



Le site de Romainville appartient à Aventis Pharma et a conservé les activités antérieures à la fusion (cf. I.1.a.ii. p.7), c'est-à-dire la recherche et le développement de médicaments agissant sur les maladies infectieuses et les maladies de l'os.

2. Le service de Documentation et d'Information Scientifiques du site de Romainville (DIS)

a. La bibliothèque scientifique

Le fonds de Romainville est géré par le logiciel Alexandria. Il comprend 19000 documents dont :

- des monographies
- des dictionnaires
- des collections
- des périodiques.

Ce fonds est consultable par l'Intranet : on peut par exemple y retrouver les sommaires des périodiques.

b. La gestion des données et des connaissances

Les connaissances internes à l'entreprise sont regroupées en différentes bases de données (informations sur les produits synthétisés à Romainville par exemple).

La base SIDoc est un acteur majeur de la gestion des connaissances : c'est en effet un système d'information scientifique donnant accès aux rapports, brevets, revues, congrès, conférences concernant les chercheurs de Romainville, ainsi qu'aux informations concurrentielles sur les projets.

c. La veille scientifique et concurrentielle

Dirigé par Mme Dominique Bidet, le service de veille scientifique et concurrentielle se compose de treize documentalistes, spécialisés chacun dans un domaine, et répartis en deux sections (cf. Annexe 1 p.39) :

- centres d'expertises (chimie, physique, pharmacologie, ...)
- projets (chimie ou biologie concernant les maladies infectieuses ou les maladies de l'os).

Les documentalistes proposent plusieurs services aux chercheurs de Romainville :

- recherches d'informations en ligne :
 - serveurs spécialisés (Dialog-DataStar, Questel-Orbit, STN)
 - Internet

- émission de profils généraux
Une veille de la littérature et des brevets est effectuée en fonction de profils particuliers (ex : Biotechnologies / Enzymologie). Les résultats sont envoyés périodiquement, généralement sous forme papier, aux personnes inscrites dans chaque profil.
- émission de profils bibliographiques
Ces profils sont individuels et informatisés. La bibliographie concernant les centres d'intérêt du client est mise à jour périodiquement par le documentaliste.
- outils "end-users" [*]
Les chercheurs peuvent rechercher seuls des références grâce à des accès électroniques à certaines bases de données, proposés par le service.

3. Objectifs du stage au sein de l'entreprise

Le service de documentation de Romainville est déjà familiarisé avec l'utilisation d'outils de recherche perfectionnés. Le serveur STN, par exemple, permet d'interroger simultanément plusieurs bases de données (Chemical Abstracts, Biosis, ...), avec, entre autres, une option de dédoublement des résultats. La requête peut même se présenter sous forme d'une structure moléculaire.

Ce serveur propose par exemple de créer une Diffusion Sélective d'Informations (DSI) : le documentaliste enregistre une requête, et choisit la fréquence à laquelle il recevra, par courrier (papier) ou par e-mail, les nouvelles références recueillies sur les différentes bases de données.

Le service a par ailleurs entrepris de mettre en place une autre DSI (cf. ci-dessus) à partir du logiciel Alexandria (cf. I.2.a. p.9), grâce à une stagiaire. Chaque employé du site pourra choisir ses centres d'intérêts parmi une liste de descripteurs, et recevra automatiquement un e-mail l'informant qu'un ouvrage susceptible de l'intéresser est arrivé à la bibliothèque.

Une DSI électronique (telle celle du logiciel Alexandria) s'assimile à la notion de "push", opposée à celle de "pull". Lors d'un **pull**, l'utilisateur est **actif** : c'est lui qui va chercher des informations. C'est le cas par exemple d'une recherche documentaire. A l'inverse, pour le **push**, l'utilisateur est **passif** : il reçoit automatiquement les nouvelles informations issues d'une source qu'il a choisie. On peut comparer le push à l'abonnement d'un périodique que l'on reçoit, même si aucune des informations qu'il contient ne nous intéresse. Ces termes sont cependant liés aux nouvelles technologies.

En revanche, les documentalistes du site de Romainville utilisent peu Internet pour rechercher des informations. Pour développer cet aspect, M. Jean-Marc Girodeau, responsable de la "gestion des données et des connaissances" (cf. I.2.b. p.9) et de l'équipe des documentalistes en charge des centres d'expertises (cf. Annexe 1 p.39), souhaite acquérir un "agent intelligent". L'objectif final est de pouvoir mettre en place grâce à cet agent une application "push". Ainsi, les documentalistes et chercheurs pourront recevoir par e-mail, automatiquement et périodiquement, des nouveautés issues d'Internet ou de l'Intranet.

L'atteinte de cet objectif passe par différentes étapes :

- définition de ce qu'est un agent intelligent
- recensement des besoins des documentalistes et de "end-users"
- élaboration d'une grille d'évaluation des agents
- recensement des agents du marché pouvant répondre aux besoins
- test des agents
- établissement d'une liste des agents répondant aux différents besoins exprimés
- installation de l'agent choisi.

II. Agents intelligents : définition

Selon le petit Larousse illustré, un "agent" est : "Celui qui est chargé de gérer, d'administrer les affaires de l'Etat, d'une société, d'un particulier." Cependant, la consultation de nombreux sites et ouvrages concernant les agents et la veille [1-15 ; 46-51] m'a amenée à constater que le concept d'"agent intelligent" est aujourd'hui très flou, ou plutôt devrais-je dire ambigu. En effet, il y a un décalage entre les travaux réalisés par les scientifiques, et les outils commercialisés sous cette appellation.

1. En théorie : le point de vue des scientifiques

Pour les scientifiques, la notion d'agent intelligent est liée à l'intelligence artificielle, et non à Internet. En effet, en 1966, c'est-à-dire avant le développement d'Internet, un premier agent intelligent fut créé : Eliza [49]. Les agents intelligents ont ensuite évolué en même temps que l'intelligence artificielle.

Aujourd'hui, les définitions sont multiples, mais plusieurs caractéristiques, dites "primaires", émergent et semblent indispensables pour pouvoir qualifier un agent d'"intelligent" :

- **l'autonomie**
L'agent doit pouvoir fonctionner automatiquement (par exemple, se connecter au web automatiquement pour y effectuer les tâches programmées), sans intervention de l'utilisateur final. Il doit pouvoir auto-définir ses propres objectifs.
- **la capacité à communiquer et à coopérer**
L'agent doit pouvoir échanger des informations avec d'autres agents, des serveurs, ou des humains. Il doit pouvoir collaborer avec ces acteurs afin d'accomplir des tâches.
- **la capacité à raisonner et à réagir à l'environnement**
L'environnement est constitué d'autres agents, Internet, Intranet, les utilisateurs, le système d'exploitation de l'ordinateur, ... L'agent doit pouvoir s'adapter aux évolutions de cet environnement, par son analyse. Il doit pouvoir enrichir son comportement avec l'expérience, en utilisant des techniques d'apprentissage.

D'après Carlo Revelli [49], seul un "vrai agent intelligent" possède ces trois caractéristiques. Des outils n'en rassemblant que deux sont des "agents mobiles", des "interfaces intelligentes", ou des "agents de gestion de profils" (selon les deux caractéristiques qu'ils regroupent).

A ces caractéristiques primaires s'ajoutent quelques caractéristiques secondaires telles que la mobilité (fonctionnement sur plusieurs systèmes et architectures), la possession d'un "caractère" propre, et même la capacité à se reproduire !

Si un agent possédait toutes ces caractéristiques, il serait en fait bien proche d'un humain. Il n'en existe pas aujourd'hui mais peut-être verrons-nous apparaître dans quelques années ces merveilles de technologie.

2. En pratique : les outils du marché

Pour les scientifiques, un agent intelligent est donc un produit de l'intelligence artificielle, devant posséder certaines caractéristiques (même si aujourd'hui aucun outil les possédant toutes n'existe).

A l'inverse, les agents intelligents du marché sont nés de l'évolution des outils de recherche sur Internet / Intranet : ils correspondent à la dernière génération de ces outils. A la fonction de recherche s'en ajoute d'ailleurs d'autres, comme la veille.

Pour bien comprendre les apports de ces agents, il est intéressant de connaître un peu les fonctionnalités de leurs "précurseurs".

a. Les précurseurs

i. Les annuaires

Aussi appelés "**répertoires**", ou "**index**", les **annuaires** sont créés et mis à jour intellectuellement. Ils regroupent un ensemble de sites sélectionnés et commentés. Ces sites sont classés par catégories, sous-catégories, ...

Yahoo! [23], un des plus célèbres annuaires, répertoriait 85000 sites en décembre 1999. Par exemple, pour obtenir une liste de sites dans l'industrie pharmaceutique, on doit choisir la catégorie "santé", puis "pharmacie", puis "sociétés pharmaceutiques@". Ce mode de recherche est très utile si on souhaite explorer un domaine inconnu.

La plupart des annuaires proposent aussi une recherche par mots-clés. Cependant, seuls quelques sites du répertoire seront donnés en réponse. En effet, cette recherche s'effectue uniquement dans l'annuaire, et seulement sur les sites indexés. Or les répertoires n'indexent pas la totalité de leurs sites.

Afin de compléter cette recherche par mots-clés, les annuaires font souvent appel à un **moteur de recherche**. C'est le cas de Yahoo! qui envoie la requête de l'utilisateur à Google [22].

ii. Les moteurs de recherche

Contrairement aux annuaires, les **moteurs de recherche** sont des outils automatiques (et non intellectuels), utiles pour des recherches précises avec des mots-clés.

Un moteur de recherche se compose de trois éléments :

- un **robot** (ou "spider", "wanderer", "crawler", ...)
Il parcourt sans cesse le Web et capture des informations contenues dans des pages.
- un **système d'indexation**
Les informations trouvées par le robot sont indexées dans le but d'être retrouvées par des mots-clés.
- un **serveur**
C'est lui qu'on interroge lorsqu'on lance une requête. Il compare les mots-clés demandés à ceux de son index et donne la liste des sites répondant à la requête.

Altavista [21] est un des moteurs les plus performants. Cependant, comme tous les autres moteurs, il n'indexe qu'une petite partie du Web. Pour augmenter le nombre de réponses à une requête, on utilise donc des **métamoteurs**.

iii. Les métamoteurs de recherche

Les **métamoteurs de recherche** permettent d'interroger simultanément plusieurs annuaires et/ou moteurs. On peut ainsi augmenter le pourcentage du Web indexé, et donc obtenir plus de réponses à une requête. Par contre, on risque d'avoir plus de bruit, d'être confronté à des lenteurs de fonctionnement, et surtout de ne pas pouvoir utiliser les options de recherche avancée de chaque moteur.

Certains métamoteurs se sont perfectionnés et offrent ainsi de nouvelles fonctionnalités :

- définition **off-line** [*] de la stratégie de recherche (requête, outils à interroger, ...)
Certains métamoteurs sont en fait des logiciels que l'on installe sur son propre ordinateur. Ces logiciels ne se connectent sur Internet / Intranet que pour effectuer la recherche.
- **traitement** des résultats d'une requête

Choix d'un agent intelligent et mise en place d'une application push.

Il est souvent possible d'éliminer les doublons et les liens non valides. Parfois, le logiciel permet aussi de regrouper les résultats par thèmes (clustérisation).

- recherche **automatisée** avec **alerte**

La requête est lancée périodiquement et automatiquement. L'utilisateur est averti, par e-mail par exemple, dès qu'une nouveauté apparaît.

Il est à noter que les annuaires et moteurs interrogés par le métamoteur peuvent être **généralistes** ou **sectoriels**. Les outils dits "sectoriels" sont spécialisés dans un domaine particulier : sciences et techniques, finance, ...

Les agents intelligents fonctionnent le plus souvent comme un métamoteur de recherche. Ils sont cependant plus performants, possèdent des fonctionnalités supplémentaires et un plus grand choix d'options.

b. Agents intelligents

Comme nous l'avons vu précédemment (cf. II. p.11), il n'existe pas de définition universelle de ce qu'est un agent intelligent. Le problème se retrouve avec les agents eux-mêmes. En effet, aucune source d'information ne s'accorde sur les différents types d'agents intelligents qui existent.

Par exemple, Sandrine Royer [50], étudiante du DESS l'an passé, et qui a approfondi le sujet, regroupe les **agents intelligents de recherche d'information et de cartographie** sur Internet en cinq catégories :

- les *aspirateurs de sites*
- les *méta-moteurs "évolués"*
- les *"Super" méta-moteurs*
- les *logiciels de cartographie des résultats*
- les *logiciels de Veille économique et de partage des connaissances.*

Carlo Revelli [49] définit quant à lui deux types d'agents :

- les **agents pour la recherche d'informations** (qui semblent regrouper les *aspirateurs*, les *méta-moteurs*, les *logiciels de cartographie* et les *logiciels de partage des connaissances* de Sandrine Royer)
- les **agents pour la veille** (qui sont en fait les *logiciels de Veille économique* évoqués par Sandrine Royer).

Afin de pouvoir classer les différents agents intelligents commercialisés, j'ai donc choisi de définir quatre catégories, correspondant aux quatre fonctionnalités que l'on retrouve chez les agents.

i. Recherche d'informations

Cette première fonctionnalité se décompose en deux parties :

- Les **sources** interrogées.

Un agent peut, tel un métamoteur de recherche (cf. II.2.a.iii. p.13), interroger des annuaires, des moteurs de recherche, des bases de données, généralistes ou sectoriels.

Il est parfois possible de rajouter des sources, en libre accès ou payantes. Dans ce dernier cas, l'agent utilise le login et mot de passe donnés par l'utilisateur, ou permet à la source de reconnaître l'adresse IP [*].

Le nombre de sources interrogées simultanément peut largement dépasser 500.

- La **requête**.

Pour aider l'utilisateur dans sa requête, ou augmenter la pertinence des informations recueillies, un agent peut posséder des fonctionnalités particulières :

- un **correcteur** d'orthographe
- un **thésaurus** (afin de définir des mots associés aux termes de la requête)
- la possibilité de choisir la **bonne signification** d'un mot (ex : avocat (juriste ou fruit))
- la possibilité d'**affiner** la requête
- le **multilinguisme** (traduction automatique de la requête en plusieurs langues)

Choix d'un agent intelligent et mise en place d'une application push.

La plupart des agents permettent une recherche en utilisant des opérateurs booléens (ET, OU), de proximité (NEAR), ou d'exclusion (SAUF). L'utilisation des parenthèses, guillemets, troncatures, ... est variable.

A ces deux groupes d'options s'en ajoutent d'autres, plus ou moins intéressantes :

- exploration des liens à partir d'une URL [*] donnée (dans le but de trouver des pages ignorées par les moteurs)
- historique des recherches avec leurs paramètres
- création de divers agents de recherche personnalisés
- mémorisation des habitudes de l'utilisateur
- lancement de plusieurs requêtes simultanément
- arrêt de la recherche avant qu'elle ne soit terminée.

ii. Aspiration de sites

L'aspiration correspond à une copie de pages web sur un ordinateur en local. Elle permet ainsi une consultation off-line de ces pages (d'où une diminution du temps de connexion, et une augmentation de la vitesse de navigation).

Ce "téléchargement" peut proposer plusieurs options :

- choix de la **profondeur**
On appelle "profondeur" le niveau d'une page web sur un site : le niveau 0 correspond à la page à partir de laquelle on commence le téléchargement.
Par exemple, en choisissant une profondeur de 1 et en commençant l'aspiration à partir de la page <http://www.pharmalicensing.com/>, on récupérera les pages :
 - <http://www.pharmalicensing.com/news>
 - <http://www.pharmalicensing.com/events>
 - etc...
- Il faut choisir une profondeur suffisamment importante si on souhaite copier un site en entier.
- limitation de l'**espace-disque** utilisé
En effet, un site copié occupe beaucoup de place sur le disque dur. Il est donc conseillé de demander à interrompre l'aspiration quand un certain espace est utilisé.
- respect ou non de la présence des **frames** [*]
- importation de fichiers **bookmarks** [*]
- lancement de plusieurs aspirations simultanément.

Que l'utilisateur automatise les téléchargements, ou qu'il les lance manuellement, il possède toujours en local la dernière version de chaque page web.

iii. Traitement des informations recueillies

Après avoir effectué une recherche, ou aspiré des pages web, l'utilisateur se retrouve avec un corpus d'informations brutes, qu'il est possible de remanier. J'ai choisi de décomposer le "traitement" en cinq parties.

- **Filtrage** des résultats.
Cette opération permet d'éliminer une partie des informations recueillies mais indésirables, c'est-à-dire d'augmenter la pertinence des résultats. Il y a plusieurs possibilités :
 - élimination des doublons
 - élimination des liens non valides
 - élimination de certaines URL, de certains noms de domaines [*]
 - élimination de certains formats (image, son, vidéo, ...)
 - élimination des documents ne contenant pas un mot-clé prédéfini.

- **Présentation** des résultats.

L'utilisateur peut souvent la personnaliser en choisissant les indications qu'il souhaite connaître pour chacun des résultats :

- titre
- description (souvent la première phrase)
- URL
- moteur de recherche utilisé
- taux de pertinence, etc...

- **Gestion** des résultats.

Cet aspect est classique mais néanmoins indispensable. Pour pouvoir réutiliser ultérieurement les informations recueillies, il est pratique de pouvoir les sauvegarder, sous forme de bases de données (en format HTML, PDF, ...) ou de fichiers HTML de bookmarks.

- **Consultation** des résultats et **recherche**.

L'utilisateur peut souvent consulter les premiers résultats alors que la recherche n'est pas terminée. Une recherche parmi les résultats (ou dans une base de données construite à partir des ces résultats) permet d'accéder directement à une information précise. Les options sont généralement les mêmes que pour la requête initiale.

- **Analyse** des résultats.

Afin de donner de la valeur ajoutée aux informations recueillies, l'agent permet parfois d'en faire une analyse. Elle peut être de deux types :

- **sémantique et statistique**

L'agent, par exemple, indexe chaque résultat dans le but de pouvoir le retrouver plus facilement ensuite. Il peut aussi générer des résumés, par extraction de passages pertinents, ou en analysant statistiquement le contenu du document. Une autre option intéressante est la clustérisation [*], qui permet par exemple de mettre en évidence des associations de mots-clés.

- **cartographique**

Dans ce cas, l'agent construit des "cartes" de relation entre les différents sites, ou entre les termes présents dans les documents.

iv. Veille

Par le terme de "**veille**", on parle de "surveillance". L'AFNOR (Association Française de NORmalisation), par exemple, la définit en 1998 comme une : "activité **continue** et en grande partie **itérative** visant à une **surveillance** active de l'environnement technologique, commercial, etc., pour en anticiper les évolutions." [4]

Ainsi, la fonction de veille des agents correspond en fait à une **automatisation** de tâches répétitives :

- recherches
- mise à jour des sources utilisées
- mise à jour des résultats
- la veille de sites :
 - téléchargement régulier ou surveillance des modifications
 - alerte (e-mail ou autre) dès que le site a changé, qu'un nouveau mot est apparu, etc...
 - mise en évidence des modifications.

Bien que l'on parle ici de téléchargement, c'est-à-dire d'aspiration (cf. II.2.b.ii. p.15), j'ai choisi de séparer cette fonctionnalité de celle de veille. En effet, certains aspirateurs n'offrent aucune autre option que celle d'avoir une copie d'un site en local.

Cette fonction induit de pouvoir choisir la **périodicité** à laquelle la veille sera effectuée. Les options de **push** (cf. I.3. p.10) entrent aussi dans cette catégorie.

III. Recensement des besoins

Afin de pouvoir choisir l'outil le plus utile pour le service de documentation de Romainville, il m'a fallu rencontrer les futurs utilisateurs de l'agent :

- Mme Dominique Bidet, responsable de la DIS et chef du service de veille scientifique et concurrentielle (cf. I.2. p.9)
- les treize documentalistes de ce service (cf. Annexe 1 p.39)
- Mme Monique Kimpflin, responsable de la bibliothèque scientifique.

L'agent pouvant aussi être utilisé par d'autres personnes du site (end-users), j'ai contacté plusieurs services : Biochimie de l'os, Coordination scientifique, Pharmacologie clinique, etc...

L'ensemble des besoins a été répertorié dans un cahier des charges fonctionnel (cf. Annexe 2 p.40).

1. Les besoins du service de Documentation et d'Information Scientifiques

Bien que chacune des personnes interrogées travaille dans un domaine précis et souvent différent des autres, les besoins sont relativement communs. Ils se regroupent en deux catégories : la **recherche d'informations**, et la **veille**.

a. Recherche d'informations

Le premier aspect de cette fonction concerne les **sources d'informations**. Les documentalistes souhaitent pouvoir faire des recherches sur :

- l'Intranet de l'entreprise
- des forums de discussion.

Il serait aussi très utile de pouvoir **rajouter** des sources, comme des moteurs de recherche. Le besoin a surtout été exprimé par rapport aux périodiques on-line [*] auxquels le service est abonné. Il y a deux possibilités pour rechercher un article :

- en interrogeant un moteur interne au site (opération récurrente)
- en regardant le sommaire (opération fastidieuse comprenant un risque de silence).

Pour faciliter les recherches effectuées sur ces périodiques, il serait donc intéressant d'avoir un agent qui puisse :

- interroger le moteur du périodique (ce qui permettrait de lancer une requête sur plusieurs de ces moteurs simultanément, et non de les interroger successivement)
- faire des recherches sur des documents en format PDF (c'est-à-dire le format de la plupart des articles on-line).

Le deuxième aspect de la recherche d'informations est la **pertinence** des résultats. En effet, Internet a été peu utilisé jusqu'à maintenant dans le service, à cause du temps relativement long passé à trouver l'information désirée (trop de bruit en général). Les documentalistes souhaitent que cette pertinence soit augmentée par :

- la possibilité de poser des requêtes poussées (booléens, parenthèses, ...)

Choix d'un agent intelligent et mise en place d'une application push.

- le choix du sens adapté d'un mot (ex : la perméabilité des sols est différente de la perméabilité intestinale)
- l'élimination des doublons
- l'utilisation de filtres efficaces.

b. Veille

Cette fonctionnalité est considérée comme la plus importante par les documentalistes. L'avantage réside surtout dans l'automatisation de tâches routinières (choix d'une périodicité d'exécution des tâches) :

- **requêtes**
L'agent interroge régulièrement des outils de recherche et envoie les nouveaux résultats trouvés par e-mail.
- **surveillance de sites**
L'agent peut effectuer cette surveillance de deux façons :
 - en aspirant (cf. II.2.b.ii. p.15) régulièrement un site (ce qui permet d'avoir toujours la version la plus récente de ce site)
 - en visitant régulièrement un site et en alertant l'utilisateur par e-mail si un changement "pertinent" est détecté.Pour l'une comme pour l'autre façon, il est utile que les modifications soient mises en valeur (fonds jaune, surbrillance, ...) : il serait dommage de perdre du temps, gagné grâce à l'automatisation, à rechercher les changements dans un site.

L'application **push** (cf. I.3. p.10), réalisée en partie pour les nouveautés d'une requête ou d'un site (envoi par e-mail), peut être étendue aux mises à jour de l'Intranet, et à des tables de matières de périodiques.

2. Les besoins des "end-users"

Les personnes que j'ai rencontrées utilisent peu Internet pour rechercher des informations, essentiellement par manque de temps. Par contre, la fonctionnalité de **veille** d'un agent correspond beaucoup plus à leurs attentes. En effet, ils pourraient ainsi obtenir des informations utiles sans perdre de temps, grâce à des requêtes et une veille de sites automatiques, complétée par un envoi des nouveautés par e-mail.

IV. Evaluation

1. La grille d'évaluation

Afin de pouvoir comparer les avantages et inconvénients de chaque agent, j'ai construit au préalable une **grille d'évaluation**, en utilisant des analyses d'agents déjà réalisées, ou des descriptions des fonctionnalités des agents :

- site de Cybion [7]
- site de Pierre Nygren [11]
- mémoire de Pierre-Alain Le Cheviller [46]
- périodique NetSources [48]
- ouvrage de Carlo Revelli [49]
- rapport de stage de Sandrine Royer [50]

Toutes ces sources m'ont permis d'avoir une idée sur ce qu'un agent intelligent était capable de faire, et donc d'établir une liste de **critères à évaluer**, notamment pour la recherche d'informations et la veille.

Choix d'un agent intelligent et mise en place d'une application push.

La grille suivante liste les critères pour les fonctionnalités correspondant aux besoins recensés à Romainville (cf. III. p.18) : **veille**, **aspiration** et **recherche d'informations**.

		OUI	NON		
V E I L L E	choix de la périodicité				
	automatisation de tâches répétitives	recherche (alerte si de nouveaux résultats sont trouvés)			
		mise à jour des sources utilisées			
		mise à jour de la base des résultats			
		autre(s) tâche(s) répétitive(s)			
	veille de sites	téléchargement régulier			
		surveillance des mises à jour ou modifications			
		alerte dès	mot-clé trouvé		
			changement dans un paragraphe précis		
			changement d'une image		
			autre(s) modification(s)		
		mise en évidence des modifications d'un site			
	modification des pages en local				
autre(s) option(s) de veille de sites					
push	on-line				
	off-line				

		OUI	NON
A S	choix de la profondeur	pages du sites	
	d'aspiration pour	sites externes	
P	choix de limitation de l'espace-disque utilisé		
I	respect de la présence de frames		
R	récupération de sommaires		
A	lancement de plusieurs aspirations simultanément		
T	importation de fichiers bookmarks		
I	données concernant le téléchargement		
O	indication des problèmes rencontrés lors de l'aspiration		
N	autre(s) caractéristique(s) de l'aspiration		

		OUI	NON
R E C H E R	diversité des sources d'information interrogées	choix de la langue des outils à interroger	
		annuaires	
		moteurs de recherche	
		bases de données	
	sources accessibles par	un mot de passe	
		une adresse IP	

C H E		autre(s) source(s)			
	possibilité d'ajout de sources	choix de la langue des nouveaux outils à interroger			
		annuaires			
		moteurs de recherche			
		bases de données			
		sources	un mot de passe		
		accessibles par	une adresse IP		
		autre(s) source(s)			
	format des documents recherchés	HTML			
		bureautique Microsoft (Word, Excel, ...)			
		PDF			
		image			
		autre(s)			
	utilisation de métamoteurs	<10 moteurs interrogés			
		10-100 moteurs interrogés			
		100-500 moteurs interrogés			
		>500 moteurs interrogés			
	requête	correcteur d'orthographe			
		langage naturel			
		multilingue (traduction de la requête en plusieurs langues)			
		possibilité de correction de la traduction automatique			
		présence d'un thésaurus			
		possibilité de choix de la bonne signification d'un mot			
		affinement de la requête			
		on-line			
		off-line			

				OUI	NON	
R E C H E R C H E	syntaxe de recherche	utilisation d'opérateurs	booléens	ET / AND		
				OU / OR		
			proximité	NEAR		
			exclusion	SAUF / NOT / BUT NOT		
		termes dans l'ordre	FOLLOWED OF ou BY			
		utilisation de parenthèses				
		utilisation de guillemets				
		troncature antérieure				
		troncature postérieure				
		différence singulier / pluriel				
		différence majuscule / minuscule				
		nécessité des accents				
		autre(s) caractéristique(s) de la syntaxe de recherche				
	vitesse de la recherche	lente				
		rapide				
		exploration des liens à partir d'une URL donnée				
		indication pour un site indisponible ou surchargé				
	mémorisation des dernières pages visitées					

sauvegarde d'une recherche (des paramètres d'une recherche)		
historique des recherches		
création de divers agents de recherche personnalisés		
mémorisation des habitudes de l'utilisateur / apprentissage		
lancement de plusieurs recherches simultanément		
arrêt de la recherche avant qu'elle ne soit terminée		
autre(s) caractéristique(s) de la recherche		

La grille complète se trouve en Annexe 3 p.43.

L'évaluation peut être différente selon le critère. Il y a deux cas :

- l'option est **présente ou non** (deux cases) :
 - case OUI cochée
 - case NON cochée
- l'option est **plus ou moins satisfaisante** (case unique) :
 - 0 : l'option est inexistante
 - + : le fonctionnement est peu satisfaisant
 - ++ : le fonctionnement est moyennement satisfaisant
 - +++ : le fonctionnement est très satisfaisant.

2. Les outils à tester

a. Les contraintes

Pour sélectionner les outils à tester, plusieurs contraintes concernant l'agent intelligent à installer ont été définies.

- **Prix de l'agent**
Le budget alloué à la mise en place de l'outil n'est pas prédéfini. L'agent intelligent que je conseillerai d'acquérir sera installé au service de documentation si son prix est en adéquation avec les services qu'il rendra.
- **Nombre d'agents à installer**
Un seul agent sera choisi. Il pourra être de deux types :
 - une "solution jetable", c'est-à-dire un outil facile à utiliser, bon marché, qui confirmera les besoins et leurs réponses
 - une "solution modulable", c'est-à-dire un outil de base sur lequel se grefferont ensuite des modules complémentaires.
- **Installation de l'agent**
Plusieurs cas sont possibles :
 - agent utilisable à partir d'un site
 - agent installé en réseau
 - agent télédiffusible
 - agent installé sur un poste, mais pouvant envoyer des résultats (par e-mail par exemple) à d'autres ordinateurs.
- **Caractéristiques de l'agent**
L'agent devra obligatoirement :
 - avoir été testé
 - pouvoir fonctionner avec Internet Explorer, en environnement Windows NT sous PC
 - être simple d'utilisation.

Le nombre d'agents à tester ne fait l'objet d'aucune contrainte : il est fonction du temps disponible et de mon appréciation personnelle.

b. Les outils

Dans un premier temps, il m'a fallu lister tous les outils commercialisés sous la dénomination "agent intelligent", ou pouvant être considérés comme tels. J'ai pour cela fait appel à plusieurs ressources :

- **sites et ouvrages** consultés lors de ma recherche d'informations sur les agents intelligents
- **17^{ème} Salon Professionnel de l'Information électronique et de l'Internet** (Idt/Net 2000), du 23 au 25 Mai 2000, au Palais des Congrès de Paris
- deux **listes de discussions** gérées par la société Cybion [7] :
 - liste "agents", consacrée aux outils de recherche d'informations et de veille sur Internet
 - liste "veille", axée sur les thématiques de veille sur Internet
- liste de diffusion de l'**ADBS** [2]
- lettre d'information **Net Hebdoc**, publiée par les Editions du Jade [8], qui apporte des informations sur les outils d'Internet
- **étudiants du DESS** Information Scientifique et Technique, promotion 1999-2000, qui faisaient aussi leur stage dans ce domaine.

A l'issue de ce recensement, je possédais une liste de 117 outils, avec ou non l'adresse URL du site qui les commercialise. J'ai recherché les adresses manquantes avec le moteur Google [22], en posant comme requête le nom de l'agent.

Afin de sélectionner les agents à tester, j'ai visité chacun de leur site et établi un tableau répertoriant certaines de leurs caractéristiques :

- **nom**
Ex : Strategic Finder
- **adresse URL** du site
Ex : <http://www.strategicfinder.com/indexfr.html>
- **matériel** informatique nécessaire
Exs : PC, Mac, ...
- **plate-forme** nécessaire
Exs : Windows NT, Unix, ...
- **navigateur** nécessaire
Exs : Internet Explorer, Netscape Navigator
- **évaluation**
2 cas : - pas d'évaluation (non)
- évaluation possible
Exs : illimitée, 30 jours, ...
- **fonctionnalités** :
 - **veille** (oui / non)
 - **aspiration** (oui / non)
 - **recherche** (oui / non)
 - **traitement**
Exs : analyse sémantique, analyse cartographique, ...
- **type**
2 cas : - modulaire (possibilité de rajouter des modules complémentaires)
- unique (pas de modules complémentaires)
- **prix**

Choix d'un agent intelligent et mise en place d'une application push.

Ce tableau m'a permis d'éliminer les agents inintéressants pour le site de Romainville, c'est-à-dire ne répondant pas aux besoins exprimés ou ne pouvant pas être mis en place, conformément au cahier des charges fonctionnel (cf. Annexe 2 p.40). Les outils à tester doivent donc tous :

- pouvoir fonctionner sur PC, avec une plate-forme Windows NT, et utiliser Internet Explorer
- pouvoir être évalués
- posséder au moins des fonctions de veille, concernant l'aspiration de sites ou la recherche d'informations, complétées ou non par un traitement des résultats.

Au final, il reste 22 agents à tester :

	NOM	ADRESSE URL	EVALUATION	VEILLE	ASPIRATION	RECHERCHE	TRAITEMENT	TYPE	PRIX
1	BullsEye 2	24	illimitée	*	*		filtrage, gestion	unique	gratuit
2	BullsEye 2 Pro		30 jours	*	*	*	filtrage, gestion, anal. sémantique	unique	249 \$
3	C4U	25	illimitée	*	*			unique	gratuit
4	Copernic 2000	26	illimitée			*	filtrage, gestion	unique	gratuit
	Copernic 2000 Plus		non			*	filtrage, gestion		39.95 \$
	Copernic 2000 Pro		non	*		*	filtrage, gestion		79.95 \$
5	eCatch	27	illimitée	*	*		filtrage, gestion	unique	179 F
6	Infonoia	28	illimitée	*		*		unique	gratuit
7	Karnak	29	illimitée	*		*	gestion	unique	gratuit
8	MemoWeb 3 Flash	30	illimitée		*		filtrage, gestion	unique	gratuit
	MemoWeb 3 Pro		non	*	*		filtrage, gestion		39.95 \$
9	Mind-it	31	illimitée	*			gestion	unique	gratuit
10	NetAttaché Pro	32	15 jours	*	*	*	filtrage, gestion	modul.	49.95 \$
11	Spyonit	33	illimitée	*		*	filtrage	unique	gratuit
12	Strategic Finder	34	illimitée	*		*	filtrage, gestion	unique	3 KF
13	Teleport Pro	35	40 fois	*	*		filtrage, gestion	unique	39.95 \$
14	The Informant	36	illimitée	*		*	gestion	unique	gratuit
15	Verity Information Server	37	démons.			*	filtrage, gestion	modul.	> 110 KF
	Verity Agent Server			*		*			> 120 KF
	Verity Ag. Serv. Toolkit								
16	WatchPortal4U	38	démons.	*		*	filtr., gest., anal. sémant. et carto.	modul.	350 KF
17	WebMirror	39	30 jours	*	*		filtrage, gestion	unique	25 \$
18	WebSnake	40	30 jours	*	*		filtr., anal. carto.	unique	197 F
19	Webspector	41	illimitée	*	*		filtrage, gestion	unique	299 \$
20	WebVCR	42	15 jours	*	*		filtrage, gestion	unique	39,95 \$
21	WebWhacker 2000	43	15 jours	*	*		filtrage, gestion	unique	49.95 \$
22	World Scanning	44	démons.	*		*	filtr., gest., anal. sémant. et carto.	modul.	80 KF

légende : - dans "ADRESSE URL" : Le nombre correspond à celui de la bibliographie.
- dans "EVALUATION" : démons. = démonstration
- dans "TRAITEMENT" : anal. = analyse

carto. = cartographique

filtr. = filtrage

gest. = gestion

sémant. = sémantique

- dans "TYPE" : modul. = modulaire

Certains agents retenus font exception aux conditions souhaitées :

- **pas d'évaluation** possible (ex : Copernic 2000 Plus et Pro)
En accord avec M. Girodeau, ces outils ont été conservés car ils permettent quand même d'évaluer une version moins perfectionnée. Les fonctions additionnelles seront indiquées comme "non testées" dans la grille de résultat.
- **démonstration** uniquement (ex : solutions de Verity)
Ces produits répondent à l'ensemble des besoins mais sont technologiquement trop lourds à installer puis désinstaller.
- **pas de fonctions de veille** :
 - *BullsEye 2*: testé en préambule de BullsEye 2 Pro, utilisable seulement pendant 30 jours.
 - *Copernic 2000 (et Copernic Plus)* et *MemoWeb 3 Flash* : seules versions qu'il est possible d'évaluer
 - *Verity Information Server* : indispensable à la mise en place de Verity Agent Server et Verity Agent Server Toolkit.

On remarque qu'il y a trois catégories d'agents intelligents :

- **gratuits** (hormis le coût de la connexion)
Ce sont des outils utilisables on-line.
- **bon marché** (entre 100 et 3000 F environ)
Ce sont des logiciels à télécharger ou à recevoir par CD-ROM.
- **solutions d'entreprises** (plus de 80000 F)
Installées sur le serveur de l'entreprise, elles sont utilisables par l'ensemble du personnel et font généralement appel à des technologies de pointe en veille, recherche d'informations, ...

3. Les tests

Les agents choisis font de la **veille** pour des **aspirations de sites** et/ou des **recherches**. J'ai donc rempli en priorité la grille d'évaluation concernant ces trois aspects. Cependant, pour avoir une idée des autres possibilités offertes par chaque agent, j'ai aussi évalué les options concernant le **traitement des résultats**.

Pour tester les **aspirations de sites**, j'ai choisi une collection d'adresses mises à jour quotidiennement :

- les news de Yahoo! [23] :
 - Internet : <http://fr.news.yahoo.com/32/>
 - Sciences : <http://fr.news.yahoo.com/45/>
 - Santé : <http://fr.news.yahoo.com/54/>
- des news dans l'industrie pharmaceutique, comme par exemple :
 - BioSpace.com : <http://www.biospace.com/b2/index.cfm>
 - pharmlicensing : <http://www.pharmlicensing.com/>
 - PharmiWeb : <http://www.pharmiweb.com/>

En ce qui concerne les **recherches d'informations**, j'ai utilisé une grille listant les centres d'intérêt de chaque documentaliste et end-user, que j'ai construite au cours du recensement des besoins. Ces mots-clés ou concepts m'ont permis de dresser une liste de sites et documents susceptibles d'intéresser chaque personne rencontrée. La liste de quelques mots-clés et concepts se trouve en Annexe 4 p.48.

Les options d'**ajout de sources** d'informations ont été évaluées avec le moteur de recherche offert par le "Journal of Biological Chemistry" (<http://www.jbc.org>), auquel le service de documentation de Romainville est abonné.

Un des problèmes majeurs d'une entreprise spécialisée dans un domaine est que les informations trouvées sur Internet sont le plus souvent **trop générales**. En effet, la plupart des agents intelligents interrogent des sources généralistes (tel Altavista [21]), et n'atteignent pas le "Web Invisible" ("invisible" car non retrouvé par les outils courants). Pour récupérer des informations plus spécifiques, il faut utiliser des outils dits **sectoriels**. Certains sites en listent quelques-uns, comme <http://www.thebighub.com/>, ou <http://www.beaucoup.com/>. J'ai donc aussi testé les agents qui permettent d'ajouter des sources avec plusieurs de ces outils (ex : le moteur de Newswise (<http://www.newswise.com/search-1.htm>), spécialisé en sciences, médecine, ...).

Une autre alternative à ce type de problème est l'utilisation d'un "**plug-in**" [*] sectoriel. L'agent Strategic Finder [34], par exemple, propose déjà un plug-in spécialisé dans les biotechnologies (Biognome), et un autre dans l'agro-alimentaire (Agrofinder). Ces plug-ins permettent d'interroger des centaines de bases de données spécialisées (répertoriées par les concepteurs du plug-in).

Pour connaître l'opinion d'autres utilisateurs d'agents intelligents, j'ai diffusé un message sur les listes de discussions "agents" et "veille" gérées par Cybion [7], et sur celle de l'ADBS [2]. J'ai par ailleurs contacté Mme Véronique Mesguich, membre du Conseil d'Administration de l'ADBS, responsable des services aux entreprises et plus particulièrement des relations extérieures et formations, qui pouvait me mettre en liaison avec des entreprises utilisant des agents. Elle m'a malheureusement confirmé ce que j'avais supposé après l'envoi de mes messages : ces informations restent aujourd'hui secrètes, car elles touchent à la veille. J'ai quand même réussi à récolter quelques avis, mais qui m'ont peu aidée.

V. Résultats

1. Problèmes rencontrés

Le problème majeur rencontré avec les agents intelligents est la **possibilité ou non de les évaluer**. Les outils testés possédaient une version d'essai, mais quatre cas se sont présentés :

	version incomplète (capacités limitées ou moins d'options)	version complète
utilisation limitée dans le temps	évaluation peu significative	évaluation significative
	<u>Ex</u> : WebMirror	<u>Ex</u> : NetAttaché Pro
utilisation illimitée	évaluation significative si seules les capacités sont limitées (<u>ex</u> : limitation du nombre de sites à surveiller)	agents gratuits fonctionnant on-line
	<u>Ex</u> : Strategic Finder	<u>Ex</u> : Spyonit

Les agents dont l'évaluation est la plus significative sont en fait ceux qui sont gratuits. Ces agents offrent des capacités intéressantes, mais qui bien sûr ne répondent pas à l'ensemble des besoins exprimés.

Les messages envoyés sur les listes de discussion ayant peu porté leurs fruits, je n'ai réuni que peu d'informations sur les outils testés, et notamment sur le bon fonctionnement des options inactives des versions d'essai. Il sera donc plus prudent de conseiller l'acquisition d'un agent évalué en version complète et limitée dans le temps, ou en version à capacités limitées.

J'ai de plus été confrontée à un autre problème, lié au site de Romainville. Par mesure de sécurité, le réseau de l'entreprise utilise un **serveur proxy** [*]. J'ai dû paramétrer chaque agent testé en indiquant :

- l'adresse de ce serveur,
- le login permettant de se connecter à Internet,
- le mot de passe correspondant,

qui lui permettront de se connecter automatiquement à Internet pour rechercher des informations ou surveiller des sites.

Malgré l'aide du service informatique, il m'a parfois été impossible de tester certains agents. C'est par exemple le cas de eCatch [27]. Au lieu d'avoir plusieurs champs à remplir, je devais mettre mes trois paramètres en une seule adresse :

`http://<login>:<mot de passe>@<serveur>` (les signes < et > étant inutiles)

Cette syntaxe ne m'a jamais permis de me connecter sur Internet pour aspirer un site et m'a donc empêché d'évaluer cet agent. Le problème ne se posait pas avec l'utilisation d'un modem (d'après Marianna Mikolajczak, étudiante du DESS, qui a réussi à utiliser eCatch de cette façon).

2. Choix d'un agent intelligent

Mon stage ne se terminant que fin septembre, je n'ai pas encore terminé mes tests. Je n'ai donc pas encore établi la liste des agents répondant aux besoins exprimés.

Cependant, mes premiers résultats font ressortir trois agents :

- **WatchPortal4U**
Ce logiciel est idéal pour le site de Romainville. Il répond à tous les besoins exprimés et a l'avantage de pouvoir fonctionner sur l'intranet de l'entreprise : c'est en effet le seul (hormis les solutions de Verity, dont la démonstration aura lieu dans quelques jours). Les seuls problèmes résident dans son prix élevé (mais en rapport à ses capacités) et dans l'impossibilité d'évaluer le produit (car il ferait appel à trop d'installations).
- **Strategic Finder**
Cet agent permet une veille efficace pour des recherches et des sites. Les résultats sont assez pertinents. L'avantage est surtout qu'il sera possible en septembre prochain d'acheter un plug-in (cf. IV.3. p.26) spécialisé dans l'industrie pharmaceutique.
- **BullsEye 2 Pro**
Comme Strategic Finder, il permet de faire de la veille pour des recherches ou des sites.

Après discussion avec M. Jean-Marc GIRODEAU, il semble que le service de documentation achètera d'abord un **agent bon marché**, type Strategic Finder ou BullsEye 2 Pro, afin de **confirmer les besoins exprimés et leurs solutions**. Mais l'objectif final sera d'acquérir une **solution d'entreprise** telle celle proposée par la société Arisem : WatchPortal4U.

Dès que l'agent sera choisi, je le mettrai en place au service de documentation. Je le paramètrerais avec les requêtes et sites souhaités par les différentes personnes rencontrées. Selon l'agent retenu, je formerai une ou deux personne(s) à son fonctionnement, ou ferai une formation globale pour que chaque utilisateur puisse paramétrer lui-même l'agent.

CONCLUSION

En bref, même si les définitions concernant les agents intelligents ne sont pas unanimes, il en existe beaucoup commercialisés sous ce nom. Sont-ils vraiment tous "intelligents" ? Je reste réservée sur ce point. En effet, au cours de mes tests, beaucoup d'outils se sont avérés décevants, soit par leurs capacités limitées, soit par la faible pertinence des résultats trouvés.

Cependant, bien que le concept d'agent intelligent ne soit pas si récent, nous ne sommes qu'au début de cette génération d'outils. On peut espérer que les agents commercialisés ressembleront bientôt aux "vrais agents intelligents" décrits par Carlo Revelli. C'est déjà presque le cas avec des solutions d'entreprises telles que WatchPortal4U de Arisem.

Malgré tout, il est possible aujourd'hui d'acquérir un agent relativement performant et assez bon marché, qui puisse répondre en partie aux besoins en veille d'un service de documentation comme celui du site de Romainville. Pour optimiser les résultats, il est nécessaire d'avoir la possibilité de rajouter des sources spécialisées dans le domaine de l'industrie pharmaceutique. Les outils généralistes retrouvent en effet peu d'informations scientifiques pointues.

Les avantages d'un stage comme celui-ci sont multiples. En plus de la participation à la vie d'un service de documentation scientifique, j'ai appris à conduire un projet par mes propres initiatives, tout en restant bien encadrée. De plus, je me suis considérablement perfectionnée dans l'utilisation d'Internet, qui est maintenant presque partout présent. Enfin, j'ai été confrontée au tabou existant concernant tout ce qui touche à la veille : le fait d'utiliser un agent intelligent dans une entreprise semble être confidentiel. Ce n'est pas le cas du site de Romainville, qui considère qu'utiliser un outil de ce type ne dévoile en rien ce qu'on surveille grâce à lui.

J'espère que mon étude permettra au site de Romainville de posséder bientôt un système performant de recherche d'informations et de veille, avec une diffusion des nouveautés issues d'Internet ou de l'Intranet grâce à un push.

BIBLIOGRAPHIE

SITES

NB : - tous les sites ont été visités pour la dernière fois le 28 Août 2000
- l'ensemble des adresses sera disponible sur l'Intranet de l'entreprise.

INFORMATIONS SUR LES AGENTS ET LA VEILLE

- 1) ABEILLE.ORG. - *Abeille.org : Les Carnets de l'Intelligence Compétitive*, [en ligne]. Adresse URL : <http://www.abeille.org>
- 2) ADBS. - *ADBS : l'association des professionnels de l'information et de la documentation*, [en ligne]. Adresse URL : <http://www.adbs.fr>
- 3) ANDRIEU, Olivier. - *Abondance : Le site des outils de recherche*, [en ligne]. Adresse URL : <http://www.abondance.com>
- 4) BERGERON, Pierrette. - *La veille stratégique : sa place à l'EBSI*, [en ligne]. Adresse URL : http://www.fas.umontreal.ca/EBSI/endirect/no11_1998/veille.htm
- 5) BOTSPOT. - *BotSpot : The Spot of All Bots on the Net*, [en ligne]. Adresse URL : <http://www.botspot.com/>
- 6) CREATEAM INTELLIGENCE SYSTEMS. - *CreaTeam : Spécialiste de la Veille sur l'Internet et l'Intranet*, [en ligne]. Adresse URL : <http://www.createam-is.com/>
- 7) CYBION. - *Veille.com : La communauté de l'intelligence économique en ligne*, [en ligne]. Adresse URL : <http://www.veille.com>
- 8) EDITIONS DU JADE. - *abcdoc.net*, [en ligne]. Adresse URL : <http://www.abcdoc.net/>
- 9) JANSON, Sverker. - *Agent-Based Systems*, [en ligne]. Adresse URL : <http://www.sics.se/isl/abc/survey.html>
- 10) MANNINA, Bruno. - *Automated Research SYStem*, [en ligne]. Adresse URL : <http://ms161u06.u-3mrs.fr/hom.html>
- 11) NYGREN, Pierre. - *Un outil de veille stratégique sur Internet*, [en ligne]. Adresse URL : <http://perso.club-internet.fr/nygren/index.htm>
- 12) RELIANCE. - *Reliance : Conseil Internet et Réseaux*, [en ligne]. Adresse URL : <http://veille.reliance.fr/>
- 13) STRATEGIC ROAD. - *Strategic Road*, [en ligne]. Adresse URL : <http://www.strategic-road.com/>
- 14) THE AGENT SOCIETY. - *The Agent Society*, [en ligne]. Adresse URL : <http://www.agent.org>
- 15) UMBC Laboratory of Advanced Information Technology. - *UMBC AgentWeb*, [en ligne]. Adresse URL : <http://agents.umbc.edu>

INFORMATIONS GENERALES

- 16) AB CREATION. - *Noms de Domaine : l'essentiel en 15 questions*, [en ligne]. Adresse URL : http://www.mon-domaine.com/noms_de_domaine.htm
- 17) FOREIGNWORD.COM. - *DictSearch*, [en ligne]. Adresse URL : <http://www.ctv.es/USERS/alberfon/dicsearf.htm>
- 18) FRASLIN, Jean-Jacques. - *Serveur proxy*, [en ligne]. Adresse URL : http://www.amgitweb.com/definitions/serveur_proxy.htm
- 19) LE CORRE, Olivier. - *Le Corre Web*, [en ligne]. Adresse URL : <http://lcorre.ifrance.com/lcorre/>
- 20) PUIMATO, Gérard. - *World Wide Web*, [en ligne]. Adresse URL : http://www.cndp.fr/notestech/22/NT022_1.HTM

OUTILS DE RECHERCHE

- 21) ALTAVISTA. - *Altavista : Recherche France*, [en ligne]. Adresse URL : <http://fr.altavista.com/>
- 22) GOOGLE FRANÇAIS. - *Google Français*, [en ligne]. Adresse URL : <http://www.google.com/intl/fr/>
- 23) YAHOO! FRANCE. - *Yahoo! France*, [en ligne]. Adresse URL : <http://fr.yahoo.com>

AGENTS TESTES

- 24) **BullsEye 2 et BullsEye 2 Pro**
INTELLISEEK. - *intelliseek : Infrastructure for Intelligent Portals*, [en ligne]. Adresse URL : <http://www.intelliseek.com/>
- 25) **C4U**
C4U. - *C4U : keeping you updated*, [en ligne]. Adresse URL : <http://www.c-4-u.com/>
- 26) **Copernic 2000, Copernic 2000 Plus, et Copernic 2000 Pro**
COPERNIC.COM. - *Copernic.com*, [en ligne]. Adresse URL : <http://www.copernic.com/fr/>
- 27) **eCatch**
ECATCH. - *eCatch : Centre de capture et de navigation Web*, [en ligne]. Adresse URL : <http://www.ecatch.com/accueil.htm>
- 28) **Infonia**

INFONOIA. - *Infonoia : L'intelligence du Web dans vos mains*, [en ligne]. Adresse URL : <http://www.infonoia.com/myagent/fr/baseframe.html>

29) **Karnak**

KARNAK. - *Karnak*, [en ligne]. Adresse URL : <http://www.karnak.com/>

30) **MemoWeb 3 Flash et MemoWeb 3 Pro**

GOTO SOFTWARE. - *Goto Software : Un clic d'avance*, [en ligne]. Adresse URL : <http://www.goto-software.com/fr/default.htm>

31) **Mind-it**

MIND-IT. - *Mind-it*, [en ligne]. Adresse URL : <http://mindit.netmind.com/>

32) **NetAttaché Pro**

TYMPANI DEVELOPMENT INC. - *Tympani Develoement Inc.*, [en ligne]. Adresse URL : <http://www.tympani.com/>

33) **Spyonit**

SPYONIT. - *Don't surf the web. Spyonit !*, [en ligne]. Adresse URL : <http://www.spyonit.com/Home>

34) **Strategic Finder**

DIGIMIND. - *Strategic Finder : Competitive Intelligence on the Web*, [en ligne]. Adresse URL : <http://www.strategicfinder.com/indexfr.html>

35) **Teleport Pro**

TENMAX.COM. - *Tenmax.com : Tools for a world of information*, [en ligne]. Adresse URL : <http://www.tenmax.com/>

36) **The Informant**

THE INFORMANT. - *The Informant : Your personal search agent on the internet*, [en ligne]. Adresse URL : <http://informant.dartmouth.edu/>

37) **Verity Information Server, Verity Agent Server, et Verity Agent Server Toolkit**

VERITY. - *Verity*, [en ligne]. Adresse URL : <http://www.verity.com>

38) **WatchPortal4U**

ARISEM. - *aristem : because you need sense*, [en ligne]. Adresse URL : <http://www.arisem.com/fr/>

39) **WebMirror**

MACCASOFT.COM. - *MaccaSoft.com : Internet and Multimedia Software*, [en ligne]. Adresse URL : <http://www.maccasoft.com/>

40) **WebSnake**

WSKA EDITIONS. - *Bienvenue chez WSKA : achetez sur Internet*, [en ligne]. Adresse URL : <http://www.wska.com/new-france.html> (site de vente)

41) **Webspector**

ILLUMIX SOFTWARE. - *Illumix Software*, [en ligne]. Adresse URL : <http://www.selfseek.com/>

42) **WebVCR**

NETRESULTS. - *Welcome to NetResults*, [en ligne]. Adresse URL : <http://www.netresultscorp.com/>

43) **WebWhacker 2000**

BLUESQUIRREL. - *BlueSquirrel : Personalizing and Mobilizing Your World*, [en ligne]. Adresse URL : <http://www.bluesquirrel.com/>

44) **WorldScanning**

CREATEAM INTELLIGENCE SYSTEMS. - *CreaTeam : Spécialiste de la Veille sur l'Internet et l'Intranet*, [en ligne]. Adresse URL : <http://www.createam-is.com/>

OUVRAGES

45) *Guide pratique de la veille technologique et stratégique sur Internet*. Paris : Innovation 128, 1998. 135 p.

46) LE CHEVILLER, Pierre-Alain. *Agents et recherche d'information sur Internet*. 82 p. : dactyl. Mémoire de fin d'études : Paris : Institut National des Techniques de la Documentation : 1997.

47) MURCH, Richard et JOHNSON, Tony. *Intelligent software agents*. Upper Saddle River : éd. Prentice Hall PTR, 1999. 210 p. ISBN 0-13-011021-3

48) NETSOURCES. *Connaître et bien utiliser l'Internet professionnel*. Paris : Bases Publication. ISSN 1270-0193

49) REVELLI, Carlo. *Intelligence stratégique sur Internet : Comment développer des activités de veille et d'intelligence économique sur le web. Moteurs de recherche. Réseaux d'experts. Agents intelligents*. 2^{ème} éd. Paris : Dunod, 2000. 220 p. ISBN 2-10-005154-7

50) ROYER, Sandrine. *Les Agents Intelligents de recherche d'information et de cartographie sur Internet*. 62 p. : dactyl. Mémoire de DESS : Nancy : Université Nancy I ; Centre de Recherche Pernod Ricard : 1999.

GLOSSAIRE

NB : Les définitions contenues dans ce glossaire ont été rédigées avec l'aide de quatre sites [16 ; 18-20], de documentalistes et d'informaticiens.

Adresse IP (Internet Protocole)

Adresse numérique caractéristique d'un utilisateur ou d'un site.

Ex : <http://195.42.35>

Pour une connexion sur Internet à partir d'un ordinateur particulier, cette adresse change à chaque nouvelle connexion.

Par contre, si l'utilisateur se connecte sur Internet à partir d'un ordinateur en réseau (dans une entreprise par exemple), l'adresse IP possède une partie fixe et une partie variable qui change régulièrement. C'est la partie fixe qui est reconnue pour un abonnement à un périodique on-line [*] par exemple.

Pour la plupart des sites, il existe une URL [*], plus facilement mémorisable.

Bookmark

Liste d'adresses de sites préférés d'une personne (signets, favoris).

Clustérisation

Regroupement de documents par catégories.

End-users (utilisateurs finaux)

Personnes pouvant utiliser un outil sans intermédiaire (consultation d'une base de données on-line [*], par exemple, sans faire appel à un documentaliste).

Frame

Cadre présent sur une page web (on trouve très souvent un frame à gauche d'une page web : le contenu du site y est inscrit afin de pouvoir aller directement dans la page recherchée).

Nom de domaine

Partie d'une URL [*] composée de :

- un nom facilement mémorisable (dans la plupart des cas)

Ex : *pharmiweb*

- une extension :

* internationale de 3 ou 4 lettres

Ex : *.com pour une société commerciale*

* géographique à 2 lettres

Ex : *.fr pour un site français.*

On-line

En étant connecté sur Internet / Intranet.

Off-line

Sans être connecté sur Internet / Intranet.

Choix d'un agent intelligent et mise en place d'une application push.

Plug-in

Programme qui s'ajoute sur un autre dans le but de lui offrir plus de possibilités.

Serveur proxy

Serveur agissant comme un filtre et gérant le trafic entre un réseau privé et Internet.

URL (Uniform Resource Locators)

Adresse d'un site.

Ex : <http://www.pharmiweb.com>

Cette adresse se compose de :

- un protocole de communication à utiliser

Exs : *http, gopher, ftp*

- ://

- un serveur

Ex : *www*

- un nom de domaine [*]

Ex : *pharmiweb.com*

ANNEXES

SOMMAIRE

ANNEXE 1

Les documentalistes..... 39

ANNEXE 2

Cahier des Charges Fonctionnel..... 40

ANNEXE 3

La grille d'évaluation :

- fonction de veille..... 43
- fonction d'aspiration..... 43
- fonction de recherche d'informations..... 44
- fonction de traitement des résultats..... 45

ANNEXE 4

Liste des quelques centres d'intérêt des documentalistes..... 48

ANNEXE 1

Les documentalistes

	NOM	PRENOM	DOMAINE
<i>Centres d'expertises</i>	BILARD	Catherine	Assurance qualité
			Cinétique

Choix d'un agent intelligent et mise en place d'une application push.

(Responsable : M. Jean-Marc Girodeau)			Galénique
	BROCCOLI	Michel	Bioanalyse
			Biophysique
			Chimie combinatoire
			Physique
	GIRODEAU	Jean-Marc	Chimie médicinale
			Management ITE (Innovation Thérapeutique et Enregistrements)
	HAUDIQUET	Martine	Brevets
	KOERTGE	Dominique	Histologie
			Hygiène industrielle
			Pharmacologie générale
			Toxicologie
	ROUILLE	Michèle	Assurance qualité
			Expérimentation animale
			Métrologie
	SEINDRIAC	Denyse	Développement chimique
Développement des procédés			

	NOM	PRENOM	DOMAINE
<i>Projets</i> (Responsable : M. Michel Manœuvrier)	COURTOIS	Patrick	Maladies de l'os : Biologie
			Screening
	LANDEROUIN	Astrid	Maladies de l'os : Biologie
	MANŒUVRIER	Michel	Biotechnologie
			Maladies infectieuses : Biologie
	RENAUD	Joëlle	Maladies infectieuses : Chimie
	SECONDA	Catherine	Maladies infectieuses : Chimie
SEMELLE	Marie-Christine	Maladies de l'os : Chimie	

ANNEXE 2

Aurélie GONDOUIN
DESS Information Scientifique et Technique

7 juillet 2000

CAHIER DES CHARGES FONCTIONNEL

I. Présentation générale du problème

1. Le demandeur

Le besoin a été exprimé par M. Jean-Marc Girodeau, responsable de la " Gestion des Connaissances " et de l'équipe de documentalistes des centres d'expertises du site de Romainville. Ce site est spécialisé dans la recherche et le développement de médicaments agissant sur les maladies infectieuses et les maladies de l'os. Le centre de documentation est composé de 12 documentalistes répartis en 2 sections :

- centres d'expertises
- projets.

Ancien centre de Recherche et Développement de Roussel Uclaf, puis de Hoechst Marion Roussel, ce site fait maintenant partie de Aventis. Il est basé :

102, route de Noisy
93235 ROMAINVILLE Cedex

2. Le produit et son marché

Le centre de documentation souhaite mettre en place un agent intelligent. Cet outil, fonctionnant sur Internet ou sur l'Intranet, peut avoir plusieurs caractéristiques différentes :

- recherche d'information
- aspiration de sites ou de pages web
- traitement des informations recueillies
- veille.

Les logiciels commercialisés exécutent en général une à deux tâches précise(s), ou toutes mais de façon moins performante. Ils peuvent être gratuits ou payants. On peut les télécharger à partir d'un site, ou les recevoir sous forme de disquette ou de CD-ROM.

3. Le contexte du projet et ses objectifs

Le projet s'inscrit dans une optique d'aide à la recherche d'informations sur Internet et l'Intranet, et à la diffusion de ces informations, avec ou sans traitement. A terme, le centre de documentation souhaite mettre en place une application "push" (diffusion automatique de nouvelles informations), à partir d'Internet et de l'Intranet de l'entreprise. Cette application sera sélectionnée de manière à répondre à un des besoins exprimés par les utilisateurs.

ANNEXE 2 (suite)

II. Expression fonctionnelle des besoins

1. Fonctions de service principales

a. Recherche et filtrage

i. Sources d'information

- Sites (ex : périodiques en ligne) auxquels le centre de documentation est abonné. L'agent utilise le mot de passe ou l'adresse IP entrés par l'utilisateur pour accéder aux informations du site.
- Forums de discussion.
- Intranet.
- Documents HTML, mais aussi PDF et autres formats.

ii. Augmentation de la pertinence des résultats

- Equations de recherche poussées (ex : utilisation de booléens).
- Requête avec le sens adapté du mot (ex : perméabilité des sols ≠ perméabilité intestinale).
- Elimination des doublons.
- Filtres efficaces.

Choix d'un agent intelligent et mise en place d'une application push.

b. Aspiration et veille

- Aspiration de certaines pages d'un site.
- Choix de la périodicité.
- Recherches automatiques et envoi des nouveaux résultats par e-mail.
- Veille de sites :
 - mise à jour régulière ou alerte par e-mail si changement pertinent
 - mise en évidence des modifications.
- Push :
 - dernières nouveautés sur l'Intranet
 - tables des matières de périodiques.

2. Fonctions de service complémentaires

a. Traitement des résultats

- Construction de bases avec les résultats (en format HTML, PDF, ou autre) et gestion de ces résultats.
- Mise en forme des résultats.
- Elimination des parties inintéressantes du document résultat.
- Analyse statistique et sémantique, clustérisation.

ANNEXE 2 (suite)

b. Autres fonctions

- Traducteur de site.
- Choix de la langue de l'information recherchée.
- Définition automatique de profils pour les utilisateurs (en fonctions de leurs habitudes et leurs requêtes).
- Construction d'un annuaire thématique pour l'Intranet.

3. Contraintes

- échéance
30 septembre 2000.
- budget
En adéquation avec le service rendu et les gains attendus.
A l'issue des tests, un agent intelligent sera conseillé. Il sera mis en place si le coût final est accepté.
- nombre d'agents à tester
Illimité.
Le nombre de tests est libre, en fonction du temps disponible et de l'appréciation personnelle.
- nombre d'outils à mettre en place
Un seul.
Une liste sera établie après les tests : elle indiquera chaque agent intéressant, avec le ou les besoins auxquels il répond.
Deux cas sont possibles :
 - mise en place d'une " solution jetable " (un seul agent, simple d'utilisation et bon marché, qui confirmera les besoins et leurs réponses)
 - mise en place d'un agent de base, auxquels se grefferont ensuite des modules complémentaires.
- installation de l'agent
Plusieurs possibilités :

Choix d'un agent intelligent et mise en place d'une application push.

- agent utilisable à partir d'un site
- agent installé en réseau
- agent télédiffusible
- agent installé sur un poste, mais pouvant envoyer des résultats (par e-mail par exemple) à d'autres ordinateurs.

- autres contraintes :
 - agent conseillé seulement après avoir été testé (d'où obligation de pouvoir tester l'agent)
 - agent fonctionnant avec Internet Explorer, en environnement Windows NT sous PC
 - agent simple d'utilisation

ANNEXE 3

La grille d'évaluation

Fonction de Veille

		OUI	NON		
V E I L L E	choix de la périodicité				
	automatisation de tâches répétitives	recherche (alerte si de nouveaux résultats sont trouvés)			
		mise à jour des sources utilisées			
		mise à jour de la base des résultats			
		autre(s) tâche(s) répétitive(s)			
	veille de sites	téléchargement régulier			
		surveillance des mises à jour ou modifications			
		alerte dès	mot-clé trouvé		
			changement dans un paragraphe précis		
			changement d'une image		
			autre(s) modification(s)		
		mise en évidence des modifications d'un site			
		modification des pages en local			
	autre(s) option(s) de veille de sites				
push	on-line				
	off-line				

Fonction d'aspiration

		OUI	NON
A S P I R A	choix de la profondeur d'aspiration pour	pages du sites	
		sites externes	
	choix de limitation de l'espace-disque utilisé		
	respect de la présence de frames		
	récupération de sommaires		
lancement de plusieurs aspirations simultanément			

T	importation de fichiers bookmarks	
I	données concernant le téléchargement	
O	indication des problèmes rencontrés lors de l'aspiration	
N	autre(s) caractéristique(s) de l'aspiration	

ANNEXE 3 (suite)

Fonction de recherche d'informations

			OUI	NON	
R E C H E R C H E	diversité des sources d'information interrogées	choix de la langue des outils à interroger			
		annuaires			
		moteurs de recherche			
		bases de données			
		sources accessibles par	un mot de passe		
			une adresse IP		
		autre(s) source(s)			
	possibilité d'ajout de sources	choix de la langue des nouveaux outils à interroger			
		annuaires			
		moteurs de recherche			
		bases de données			
		sources accessibles par	un mot de passe		
			une adresse IP		
		autre(s) source(s)			
	format des documents recherchés	HTML			
		bureautique Microsoft (Word, Excel, ...)			
		PDF			
		image			
		autre(s)			
	utilisation de métamoteurs	<10 moteurs interrogés			
		10-100 moteurs interrogés			
		100-500 moteurs interrogés			
		>500 moteurs interrogés			
requête	correcteur d'orthographe				
	langage naturel				
	multilingue (traduction de la requête en plusieurs langues)				
	possibilité de correction de la traduction automatique de la requête				
	présence d'un thésaurus				
	possibilité de choix de la bonne signification d'un mot				
	affinement de la requête				
	on-line				
off-line					
syntaxe de	utilisation	booléens	ET / AND		

Choix d'un agent intelligent et mise en place d'une application push.

recherche	d'opérateurs		OU / OR			
		proximité	NEAR			
		exclusion	SAUF / NOT / BUT NOT			
	termes dans l'ordre de la requête		FOLLOWED OF ou BY			
	utilisation de parenthèses					
	utilisation de guillemets					
	troncature antérieure					
	troncature postérieure					

ANNEXE 3 (suite)

Fonction de recherche d'information (suite)

			OUI	NON
R E C H E R C H E	syntaxe de recherche (suite)	différence singulier / pluriel		
		différence majuscule / minuscule		
		nécessité des accents		
		autre(s) caractéristique(s) de la syntaxe de recherche		
	vitesse de la recherche	lente		
		rapide		
	exploration des liens à partir d'une URL donnée			
	indication pour un site indisponible ou surchargé			
	mémorisation des dernières pages visitées			
	sauvegarde d'une recherche (des paramètres d'une recherche)			
	historique des recherches			
	création de divers agents de recherche personnalisés			
lancement de plusieurs recherches simultanément				
arrêt de la recherche avant qu'elle ne soit terminée				
autre(s) caractéristique(s) de la recherche				

Fonction de traitement des résultats

			OUI	NON		
T R A I T E M E N T	filtrage des résultats	élimination des doublons				
		élimination des liens non valides				
		choix d'exclusion de	certaines URL			
			certains formats			
			diverses extensions des navigateurs (Active X, ...)			
			noms de serveurs			
		noms de domaines				
		élimination des documents ne contenant pas un mot-clé prédéfini				
	choix du nombre de références à rechercher					
	autre(s) possibilité(s) de filtrage					
	présentation des résultats	indications concernant la page elle-même	titre / nom			
			description (souvent la 1ère phrase)			
URL						
chemin suivi pour y accéder (source)						

Choix d'un agent intelligent et mise en place d'une application push.

		autre(s) indication(s)		
	indications concernant la recherche de la page	requête utilisée		
		moteur de recherche utilisé		
		taux de pertinence		
		heure de lecture de la page		
		nombre de références trouvées		
		autre(s) indication(s)		

ANNEXE 3 (suite)

Fonction de traitement des résultats (suite)

				OUI	NON		
T R A I T E M E N T D E S R E S U L T A T S	gestion des résultats	sauvegarde					
		choix du format de sauvegarde					
		choix du lieu de sauvegarde					
		archivage	construction de bases de données de gestion des résultats	format HTML			
				format bureautique			
				format PDF			
				autre(s) formats			
			transformation en fichiers HTML de bookmarks	sans résumé			
				avec résumé			
		déplacement					
		copie					
		destruction / élimination					
		impression					
		export des résultats					
	envoi par e-mail						
	extraction et regroupement des adresses e-mail présentes sur un site						
	consultation et recherche	consultation des premiers résultats avant la fin de la recherche					
		annotation					
		navigation off-line					
		recherche sur les bases construites	requête	correcteur d'orthographe			
				langage naturel			
				multilingue			
				possibilité de correction de la traduction			
				présence d'un thésaurus			
				choix de la bonne signification d'un mot			
			affinement de la requête				
			syntaxe	utilisation d'opérateurs	booléens	ET / AND	
					OU / OR		
		proximité			NEAR		
		exclusion		SAUF / NOT / ...			
		recherche dans l'ordre	FOLLOWED OF				
utilisation de parenthèses							
utilisation de guillemets							
troncature antérieure							

Choix d'un agent intelligent et mise en place d'une application push.

			troncature postérieure		
			différence singulier / pluriel		
			différence majuscule / minuscule		
			nécessité des accents		
			autre(s) caractéristique(s)		

ANNEXE 3 (suite)

Fonction de traitement des résultats (suite)

				OUI	NON		
T R A I T E M E N T	consultation et recherche	recherche sur les pages téléchargées par	mots-clés				
			URL				
			format				
			autre(s) critère(s)				
	analyse sémantique et statistique	indexation sémantique	conception automatique de résumé par	extraction des concepts du document			
				extraction de mots-clés			
				ajout d'une expression issue du document pour différencier un mot à 2 significations			
		choix de la taille des résumés	construction de tables des matières	clustérisation	extraction de passages pertinents		
					analyse statistique du contenu		
					autre(s) technique(s)		
	analyse cartographique	carte de relations entre	remaniement de la carte	les sites (pages)			
				les thèmes			
				autre(s) critère(s)			

ANNEXE 4

Liste de quelques centres d'intérêt des documentalistes

NB : - les mots en italique sont la traduction en anglais des mots de la colonne de gauche : cette traduction a été effectuée par les documentalistes, ou en utilisant les dictionnaires de *foreignword.com* [17]

- les mots couvrant deux colonnes n'admettent pas de traduction utilisée couramment

- cette présente liste est incomplète : en effet, la liste utilisée pour les tests comprenait des mots-clés et concepts liés à des projets confidentiels

maladie infectieuse	<i>infectious disease</i>
antiinfectieux	<i>antiinfectious</i>
antibiotique	<i>antibiotic</i>
antimicrobien	<i>antimicrobial</i>
antifongique	<i>antifungal</i>
antibactérien	<i>antibacterial</i>
maladie de l'os	<i>bone disease</i>
ostéoporose	<i>osteoporosis</i>
nouveau produit	<i>new product</i>
nouvelle cible	<i>new target</i>
assurance-qualité	<i>quality assurance</i>
bonne pratique de fabrication	<i>good manufacturing practice</i>
bonne pratique clinique	<i>good clinical practice</i>
bonne pratique de laboratoire	<i>good laboratory practice</i>
affaire réglementaire	<i>regulatory affair</i>
norme	<i>norm / standard</i>
<i>draft</i>	
<i>guideline</i>	
Europe	<i>Europe</i>

Choix d'un agent intelligent et mise en place d'une application push.

Etats-Unis	<i>United States (USA)</i>
Japon	<i>Japan</i>