

ENSSIB

Ecole Nationale Supérieure des Sciences de
l'Information et des Bibliothèques

**Université
Claude Bernard - Lyon 1**

DESS en INGÉNIERIE DOCUMENTAIRE

Rapport de Stage

Mise en place d'un portail de veille sur Internet

Nicolas Bayon

Sous la direction de
D. Roussel EADS Aérospatiale Matra Centre Commun de Recherche Suresnes

2000

BIBLIOTHEQUE DE L'ENSSIB



8140368

R1

M 2000 ID ST 5

ENSSIB

Ecole Nationale Supérieure des Sciences de
l'Information et des Bibliothèques

**Université
Claude Bernard - Lyon 1**

DESS en INGÉNIERIE DOCUMENTAIRE



Rapport de Stage

Mise en place d'un portail de veille sur Internet

Nicolas Bayon

Sous la direction de
D. Roussel EADS Aérospatiale Matra Centre Commun de Recherche Suresnes

2000

Remerciements

J'adresse mes plus sincères remerciements à l'ensemble du service DCR/IKT du Centre Commun de Recherche EADS de Suresnes. A Monsieur David Roussel pour m'avoir offert une structure d'accueil pour mon stage, pour l'attention qu'il a portée à l'avancée du projet, pour son aide technique tout au long du projet sans laquelle il n'aurait pas pu être mené à bien. A Madame Regnier-Prost, Madame Monceaux et Monsieur Meyer pour leur accueil au sein du service. A Monsieur Jancek pour ses conseils avisés.

Mise en place d'un portail de veille sur Internet

Résumé et descripteurs

Résumé

Ce rapport de stage décrit un projet d'alimentation automatique d'un portail Web dans le cadre de la veille concurrentielle réalisée au sein de EADS Aerospatiale Matra Missiles. Les missions qui m'ont été confiées concernent la conception et la réalisation d'un démonstrateur qui repose sur l'enchaînement suivant :

- 1. En amont du portail : utilisation d'outils de collecte d'informations sur Internet (outils de recherche thématique, outils de surveillance, outils d'aspiration)*
- 2. Visualisation et contrôle éditorial des documents rapatriés via un portail*
- 3. Sélection des documents à signaler dans la base de données Missile News*
- 4. Sauvegarde des signalements des documents par catégories*

Ce stage s'est déroulé au Centre Commun de Recherche d'EADS en plusieurs étapes : Etude du sujet, découverte d'environnements de programmation, rédaction de documents de conception, présentations aux clients et installation de la solution proposée.

Descripteurs

Base de données, flux d'information, Internet, veille.

Abstract

This probation report is intended to describe the realization of a technology watch web portal for EADS Aerospatiale Matra Missiles. During this period, the realization plan was :

1. Using and comparing search tools like e-Catch, Copernic, DigOut4U
2. Editing and controlling search results in a portal
3. Selecting interesting documents to export to "Missile News" database
4. Saving document datas

This probation period took place in EADS "Centre Commun de Recherche"

Keywords

Database, information flow, Internet, technology watch.

TABLE DES MATIERES

I. PRÉSENTATION DE L'ENTREPRISE 4

 I.1. EADS 4

 I.1.1. *Présentation générale* 4

 I.1.2. *Organigramme* 5

 I.2. AEROSPATIALE MATRA 5

 I.2.1. *Historique* 6

 I.2.2. *Les sites en France* 6

 I.2.3. *Les produits* 7

 I.2.4. *Position sur le marché* 7

 I.3. LE CENTRE COMMUN DE RECHERCHE LOUIS-BLÉRIOT (CCR) 8

 I.3.1. *Présentation générale* 8

 I.3.2. *Organisation* 9

 I.3.3. *Productique Documentaire et Traitement de l'Information* 10

II. PRÉSENTATION DU PROJET VIAMM 12

 II.1. FORMULATION INITIALE : « TECHNIQUES D'ALIMENTATION D'UN PORTAIL POUR LA VEILLE » 12

 II.1.1. *Contraintes* 13

 II.1.2. *Techniques de recherche d'information adoptées* 13

 II.1.2.1. *Moteurs de recherche thématique, outils de surveillance et répliation d'information* 13

 II.1.2.2. *Optimisation des formulaires d'accès à des pages dynamiques* 14

 II.1.2.3. *Mise en place d'agents intelligents* 15

 II.1.3. *Comparaison avec les offres commerciales de solution de portail pour la veille* ... 18

 II.1.3.1. *Caractérisation générale de l'offre* 18

 II.1.3.2. *Mediapps* 19

 II.1.3.3. *QwamSystem* 20

 II.1.3.4. *Arisem* 21

 II.2. TECHNOLOGIES WEB ADOPTÉES 22

 II.2.1. *Contraintes* 22

 II.2.2. *Choix des technologies* 22

 II.2.2.1. *Le langage Perl* 22

 II.2.2.2. *Le langage PHP* 24

 II.2.2.3. *Le langage JavaScript* 24

 II.2.2.4. *La base de données MySQL* 25

 II.2.2.5. *Le serveur Apache* 25

III. CONDUITE DU PROJET 26

IV. RÉALISATION DU PORTAIL VIAMM : DE L'ILLUSTRATION HTML À LA MAQUETTE VIAMM V4 30

 IV.1. UNE ARCHITECTURE MODULAIRE 30

 IV.1.1. *Organisation du disque dur du poste* 30

 IV.1.2. *La structure de la base de données* 32

 IV.2. UNE INTERFACE SIMPLE 34

 IV.2.1. *Structure générale du démonstrateur* 34

 IV.2.2. *Ergonomie des rubriques* 35

 IV.2.2.1. *Page de présentation des collectes* 36

 IV.2.2.2. *L'accès aux moteurs de recherche en ligne* 42

 IV.2.2.3. *Les bookmarks* 44

 IV.2.2.4. *La rubrique "Outils"* 46

 IV.3. UNE GESTION SOUPLE DES FLUX DE DONNÉES 48

 IV.4. UNE DOCUMENTATION INTERACTIVE 52

V. VIAMM EN PRATIQUE 55

V.1.	SCÉNARIOS D'UTILISATION	55
V.1.1.	<i>Mises au points à réaliser avec les différents outils</i>	56
V.2.	COLLECTE DES RETOURS UTILISATEURS	57
VI.	BILAN	62
VI.1.	BILAN TECHNIQUE.....	62
VI.2.	BILAN PROJET	63
VII.	BIBLIOGRAPHIE	64
Annexes		

I. Présentation de l'entreprise

I.1. EADS

I.1.1. Présentation générale

The European Aeronautic Defence and Space Company (EADS N.V.) est la première société européenne d'aéronautique et de défense. EADS cumule les activités de ses partenaires fondateurs que sont Aérospatiale Matra S.A. (France), Construcciones Aeronáuticas S.A. (Casa-Espagne) et Daimler Chrysler Aerospace AG (Dasa-Allemagne).

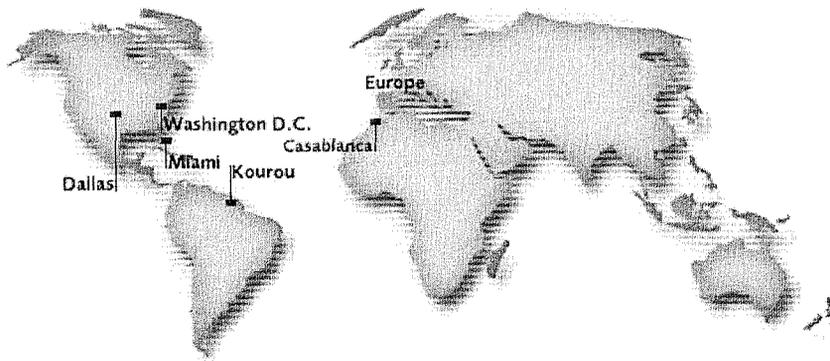


Figure 1 : *Les principaux sites EADS au niveau mondial*

I.1.2. Organigramme

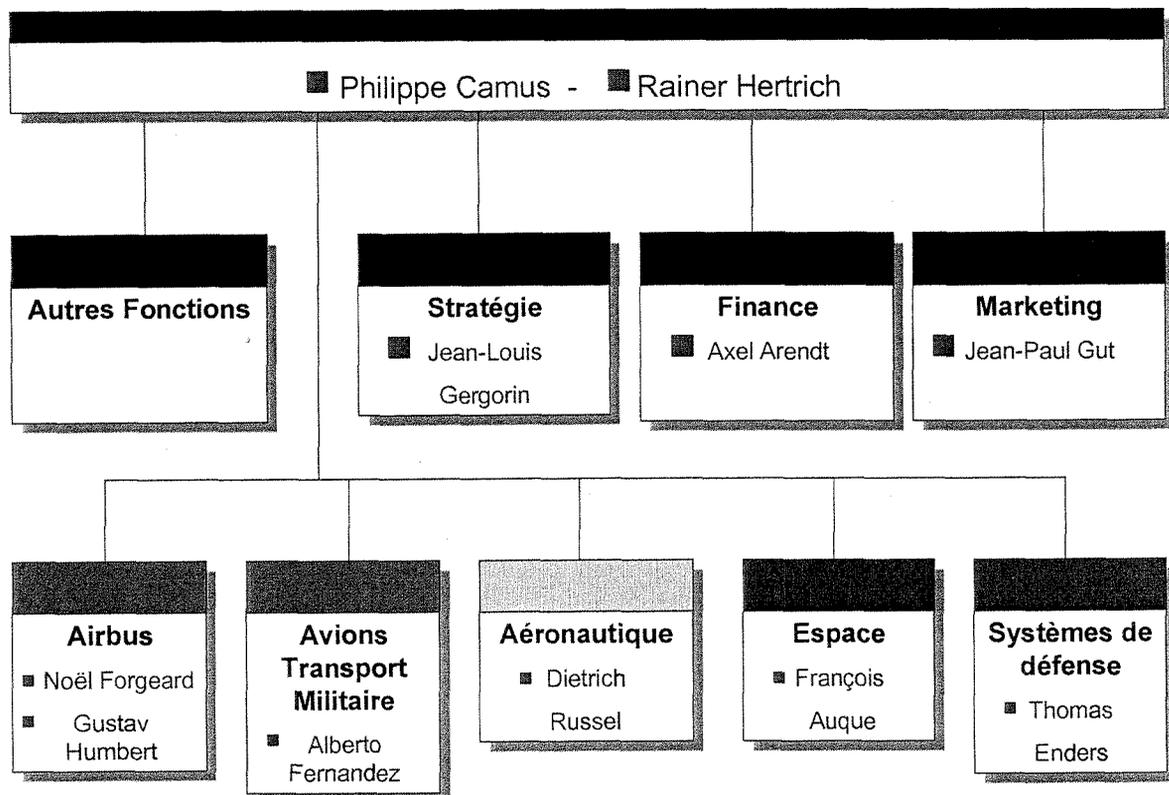


Figure 2 : Organigramme des hauts dirigeants du groupe EADS

I.2. Aérospatiale Matra

C'est la branche française du groupe EADS. Elle est elle-même le résultat d'une fusion récente entre Aérospatiale et Matra Haute Technologie (Juin 1999). Son siège social est situé 37, Bd de Montmorency 75016 Paris. Son objet social est « Etude, fabrication et commercialisation de matériels du secteur aéronautique et spatial ». La président du Directoire est Philippe Camus.

I.2.1. Historique

- 1906 : Fondation des ateliers Blériot.
- 1936 : Création des sociétés nationales d'aéronautique françaises.
- 1945 : Constitution de Matra (Mécanique Aviation TRAction).
- 1970 : De la fusion de trois sociétés Nord-Aviation, Sud-Aviation et SEREB, naît le groupe Aérospatiale.
- 1998 : Privatisation d'Aérospatiale s'accompagnant de la fusion avec Matra Hautes Technologies.
- 1999 : Introduction en bourse d'Aérospatiale Matra.

I.2.2. Les sites en France

Même si le siège social est situé à Paris, de nombreux sites se répartissent partout en France.

Siège Social : Paris

Centre de Recherches : Suresnes

Aéronautique : Toulouse, Nantes, Saint-Nazaire, Méaulte, Sogerma Bordeaux et Rochefort, Le Bourget, Tarbes

Missiles : Bourges, Châtillon, Celerg, Cilas, Matra BAe Dynamics

Eurocopter : Marignane, La Courneuve

Espace/Défense : Matra Marconi Space, Sodern, Les Mureaux, Aquitaine

Services et Télécommunications : ISTI, Matra Nortel Communication, Matra Système et Information, Matra Datavision

Figure 3 : Répartition des sites et filiales d'Aérospatiale Matra France

I.2.3. Les produits

Ils reflètent la diversité des branches de l'entreprise.

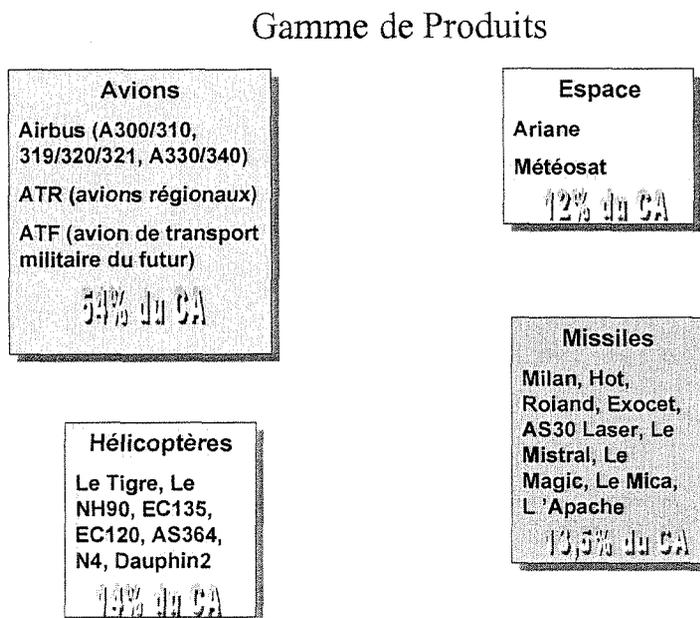


Figure 4 : Gamme des produits commercialisés par l'entreprise

Le chiffre d'affaires représente 12,3 milliards d'euros (80,6 milliards de francs). 90% de celui-ci est réalisé en coopération internationale. Un réseau de filiales et de bureaux permanents à l'étranger confirme la dimension internationale de l'entreprise.

I.2.4. Position sur le marché

N°1 Mondial dans les lanceurs commerciaux

N°1 Mondial dans les hélicoptères civils

N°1 Mondial dans les réseaux de radiocommunication numériques sécurisés

N°2 Mondial dans les avions commerciaux

N°2 Mondial dans les missiles tactiques

N°2 Mondial dans les satellites

N°2 Mondial dans les avions d'affaires (avec Dassault Aviation)

N°3 Mondial dans les avions de transport régional

N°3 Mondial dans les avions de combat (avec Dassault Aviation)

Les prises de commandes représentent 17,6 milliards d'euros (115 milliards de FF). Le résultat net consolidé est de 436 millions d'euros (2,86 milliards de FF). 80% du chiffre d'affaires sont réalisés dans des programmes européens.

Comme le laissent présager ces résultats, les enjeux sont énormes et la concurrence est extrême. C'est dans ce contexte que s'est déroulé mon projet de stage.

I.3. Le Centre Commun de Recherche Louis-Blériot (CCR)

I.3.1. Présentation générale

Ce centre, situé à Suresnes, regroupe la plus grande partie des activités de recherche de l'ensemble de la société. Créé en 1915, par Louis Blériot, il participe activement à une production massive de matériels civils et militaires notamment lors de la première guerre mondiale. Il prend part aux programmes aéronautiques Caravelle, Concorde, Airbus...

Le GIE Aérospatiale Matra CCR comprend 260 personnes localisées à Suresnes dont une quinzaine de collaborateurs répartis entre les antennes de Toulouse et Bourges. Ces effectifs se répartissent entre 60% d'ingénieurs et cadres et 40% de techniciens.

En dehors des travaux de recherche, près de 20% de son activité consiste en des missions de support technique (expertise et assistance) ou des prestations de services telles que la normalisation et la documentation.

I.3.2. Organisation

Le CCR a pour mission de promouvoir les innovations technologiques présentant un intérêt commun pour le groupe. Il se compose des différents départements techniques suivants :

- Pôle calcul scientifique : réalisation de logiciels de simulation scientifique
- Moyens communs : test et essais au service des autres départements
- Normalisation / Documentation : départements de service, rédaction et gestion de normes internes, documentation scientifique et technique, veille technologique et brevets. Un service de propriété intellectuelle assure la protection, la confidentialités des travaux effectués dans le cadre de ces grands thèmes de recherches.
- Matériaux : élaboration et caractérisation des matériaux et traitements de surface, contrôle non destructif
- Procédés Matériaux Composites : mise en œuvre des matériaux composites thermodurs et thermoplastiques
- Procédés Matériaux Métalliques : mise en œuvre des matériaux métalliques (soudage, prototypage rapide)
- Simulation des procédés : simulation de la mise en œuvre des matériaux métalliques et composites (mise en forme de tôles, emboutissage formage superplastique...)
- Support technique : réalisation d'expertises sur pièces produits
- Environnement des Systèmes : compatibilité électromagnétique, tenue à la foudre, tenue aux radiations, furtivité, composants électroniques, optoélectronique...
- Productique et Technologies de l'Information

I.3.3. Productique Documentaire et Traitement de l'Information

Mon projet de stage s'est déroulé dans le service "Productique Documentaire et Traitement de l'Information"

Les travaux menés par ce service regroupent les thèmes du traitement du document et du langage naturel, les technologies vocales et dans une moindre mesure (en collaboration avec d'autres services) de la réalité virtuelle et l'intelligence artificielle.

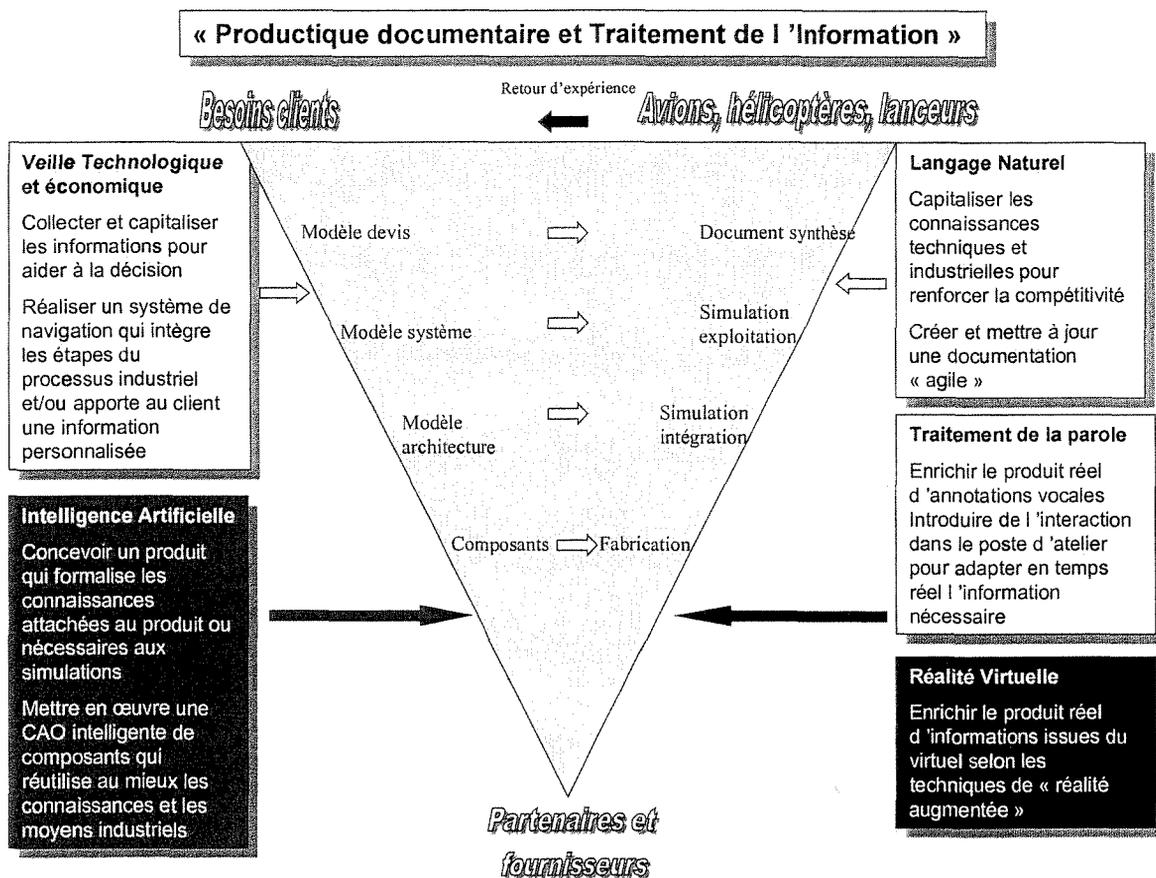


Figure 5 : Schéma présentant les différentes activités de la filière "Productique et Technologies de l'Information"

Plus précisément, la productique documentaire regroupe les thèmes terminologie, traduction multilingue, contrôle des langues (normes de rédaction de documents). Le

traitement de l'information regroupe lui, la veille technologique et économique, navigation documentaire (cartographie, résumés, ...) et diffusion sélective d'informations.

Les missions principales du service DCR/IKT sont les suivantes :

- - Caractériser et qualifier les produits permettant le contrôle du langage et la traduction linguistique des documents
-
- - Promouvoir des produits pour les systèmes de veille ainsi que des logiciels pour un exploration efficace des documents techniques
-
- - Rechercher et évaluer les technologies de reconnaissance vocale
-
- - Assister les différentes entités dans le cadre de la consultation des services proposés
-
- - Réaliser des maquettes illustrant des procédés innovants
-
-

Les experts du traitement de l'information du service ont un rôle de conseil en ce qui concerne le choix et les conditions d'utilisation des outils et des méthodes proposés. Ils contribuent à diffuser les résultats des recherches nationales et internationales dans la société. Les travaux de recherches sont réalisés dans le cadre d'études prospectives.

II. Présentation du projet VIAMM

Le projet VIAMM repose sur la fourniture au centre de documentation de Châtillon (Aérospatiale Matra Missiles) d'une solution permettant d'optimiser l'alimentation d'une base de données Lotus Notes appelée "Missile News". Cette base est prévue pour contenir des signalements d'informations disponibles sur Internet. Minimale, un signalement consiste en un titre, une date, une URL, une ou plusieurs catégorisation(s) et un résumé destiné à appréhender le contenu du document (il peut s'agir de phrases clés). Le nom VIAMM est l'acronyme de Veille Internet Aérospatiale Matra Missiles.

II.1. Formulation initiale : « Techniques d'alimentation d'un portail pour la veille »

C'est sous cette forme - technique d'alimentation d'un portail pour la veille" que le projet VIAMM a été envisagé avant même le début de sa réalisation. On verra que cette présentation du projet a subi quelques modifications au fur et à mesure du développement du portail.

Le maquette finale repose sur l'enchaînement suivant :

1. En amont du portail : utilisation d'outils de collecte d'informations sur Internet (outils de recherche thématique, outils de surveillance, outils d'aspiration) dont certains sont capables d'extraire un ensemble de phrases remarquables en fonction d'une requête
2. Visualisation et contrôle éditorial des documents via un portail
3. Sélection des documents à signaler dans la base de données Missile News
4. Sauvegarde des signalements des documents par catégories

Avant de présenter la maquette elle-même, voici une présentation des contraintes du projet puis des techniques utilisées, techniques dont l'étude approfondie a occupé une

partie de mon stage (voir [ROU&BAY 2000a], [ROU&BAY 2000b], [ROU&BAY 2000c]).

II.1.1. Contraintes

Elles sont au nombre de quatre. Premièrement, les outils utilisés pour la conception doivent être disponibles et de faible coût afin de faciliter le déploiement.

Deuxièmement, l'architecture des postes de travail présentes chez le client est spécifique : le réseau intranet est totalement séparé du réseau Internet. Actuellement, une bi-connexion est à l'étude mais la consultation d'Internet se fait toujours sur des postes indépendants montés en un réseau local. Ces postes fonctionnent sous Windows NT.

Troisièmement, les informations doivent pouvoir être signalées sur l'intranet. Ces signalement sont à intégrer dans une base Lotus Notes appelée "Missiles News".

Enfin, les utilisateurs finaux ne sont pas spécialistes d'Internet et souhaitent une grande facilité d'emploi, une aide en ligne, et des possibilités de récupération des erreurs...

Autant d'exigences et de contraintes qui ont permis d'orienter le choix des technologies.

II.1.2. Techniques de recherche d'information adoptées

II.1.2.1. Moteurs de recherche thématique, outils de surveillance et réplique d'information

Les outils situés en amont du portail VIAMM réalisent une recherche thématique et une aspiration des sites sélectionnés. Les outils choisis pour ce projet l'ont été pour :

- leur richesse fonctionnelle ;
- leur complémentarité ;



- leur facilité d'intégration dans l'architecture des stations concernées par l'utilisation future de VIAMM ;
- leur faible coût (la plupart sont gratuits).

Les outils retenus pour le projet sont Copernic 2000 Pro, Digout4U 1.5, e-Catch 3.0.

Une grille fonctionnelle comparative a été intégrée au portail Viamm (voir annexe I.3)

Dans un premier temps, nous avons évalué les limites de ces outils vis-à-vis de certains sites indiqué par le client (Aérospatiale Matra Missiles). Cette estimation s'est appuyée sur des outils tels que Astrasite Manager (cartographie de sites) et prend en compte des problèmes de format et de mise à jour des informations rapatriées.

Nous avons observé :

- des points faibles sur l'automatisation des tâches que nous avons compensés par l'utilisation de l'outil WinMacro ;
- des problèmes vis-à-vis des formulaires d'accès à des pages dynamiques (voir section suivante).

II.1.2.2. Optimisation des formulaires d'accès à des pages dynamiques

Il existe actuellement des sites en ligne d'information à valeur ajoutée d'accès gratuits. Deux problèmes sont posés par ces sites. Premièrement, les pages d'accueil varient. Deuxièmement, le rapatriement automatique du contenu de ces sites nécessite la mise au point de formulaires. Par exemple, lorsqu'un formulaire possède différents champs à interroger avec des mots clés, il s'agira de fixer, pour chacun de ces champs, des correspondances avec des champs d'information où se trouvent les termes demandés. Ainsi, le champ *keyword* d'une notice bibliographique entre en compétition avec d'autres champs (ex: titre, résumé, résumé en anglais, mots clés en français, ...). Si le nombre de documents est relativement important, mieux vaut choisir le champ le plus spécifique et se baser sur un thésaurus. Pour les notices bibliographiques, la pertinence des résultats sera probablement supérieure à partir d'une sélection sur le champ *titre*

puisque les auteurs ont choisis d'y inclure des mots clés en même temps que des "marqueurs d'originalité".

II.1.2.3. Mise en place d'agents intelligents

La mise en place d'agents intelligents est complémentaire des tâches précédemment évoquées. Les principaux objectifs sont de :

- récupérer des informations du Net sans intervention manuelle (après une phase de paramétrage)
- trouver des informations non indexées sur des sites à surveiller
- filtrer la masse d'informations présente sur des sites à surveiller avant rapatriement
- signaler les évolutions de certains sites

Agents "pull"

La finalité de ces agents est d'aider à surveiller un ou plusieurs sites Web et, éventuellement, les archives des forums de discussion ou des listes de diffusion. On distingue deux possibilités :

Les agents d'alerte par e-mail (MyUpdate, The Informant...) vous envoient un courrier électronique chaque fois que tout ou partie de la page Web que vous avez sélectionnée est modifiée (MyUpdate), ou bien quand une recherche d'informations produit de nouveaux résultats au sein d'un ou plusieurs moteurs (The Informant). Dans certains cas, (The Informant), il est également possible de filtrer les changements à surveiller en choisissant des mots-clés appropriés.

Les problèmes posés les agents d'alertes que nous avons pu tester sont :

- le choix des mots clés est très difficile ;
- le nombre d'URL qui peuvent être surveillées à partir de mots clés est limité ;
- toutes les activités de veille sont généralement gratuites et sont gérées par un serveur distant (vous n'avez aucun logiciel à installer sur votre ordinateur). Il s'agit donc d'outils très flexibles et faciles à utiliser mais avec lesquels on ne peut pas mener des activités de veille trop poussées (problème de confidentialité).
- Les agents d'alerte associés à des pages particulières ne réalisent pas d'aspiration paramétrable de ces pages. Il n'est pas possible par exemple d'aspirer les nouvelles pages liées.

Les agents d'alerte "off-line" (e-Catch, Copernic Pro 2000) sont des logiciels clients qui s'installent sur votre ordinateur et que vous pouvez finement personnaliser afin de réaliser des activités de veille. Ces agents réalisent les mêmes tâches que les agents on-line mais permettent également de copier sur votre ordinateur des sites entiers. Il suffit de leur indiquer l'adresse du site et l'heure à laquelle vous souhaitez qu'ils s'activent. Ils vont se connecter de façon automatique au réseau afin de rapatrier les données qui vous intéressent. Vous pouvez ensuite mener vos activités de recherche et de veille à partir des données enregistrées localement (avec un affichage immédiat). L'inconvénient majeur de ces agents est qu'ils doivent (pour des raisons d'efficacité) se baser sur les dates de dernières modifications qui sont souvent peu pertinentes. Il suffit qu'une virgule ou qu'une image soient déplacées pour que l'agent interprète qu'il y ait eu un changement.

Agents "push"

Les agents "push" sont très efficaces pour suivre l'actualité et pour diffuser des informations à des cibles très larges. Là aussi, on distingue deux cas :

Les agents "push on-line" (MyYahoo!, My NewsPage, My Excite, eBusiness Center,...) permettent d'obtenir des nouvelles personnalisées sous forme de rubriques thématiques (actualité sportive, actualité politique, actualité financière, actualité météo,...etc) ainsi que beaucoup de services annexes (réservations, achats, loisirs, communautés électroniques...). Tout fonctionne en ligne et vous n'avez pas besoin d'installer de logiciel client. Les informations sont diffusées au sein d'une page de sommaire personnelle qui s'active dès lors que vous avez défini votre profil d'utilisateur. Ce genre de service est simple et assez efficace mais très limité pour des activités de veille sophistiquées. En effet, à part les contraintes liées à la connexion Internet, vous disposez souvent d'un choix très restreint d'options pour personnaliser votre page (exception faite de MyNewsPage et de eBusiness Center).

Les agents "push off-line" reposent sur le même principe mais nécessitent d'installer un logiciel client. Ainsi, une fois que vous avez défini votre profil au sein d'une liste d'options de plus en plus riches, votre logiciel télécharge automatiquement, à partir d'un serveur distant, le contenu des chaînes auxquelles vous avez souscrit.

Les agents de recherche avancée

Les moteurs sectoriels sont des outils uniquement spécialisés dans un domaine précis (finance, assurance, éducation, littérature, etc...). Ce sont souvent des bases de données indépendantes et très riches qui ne peuvent pas être atteintes par les robots de recherche généralistes (il s'agit de ce que les spécialistes appellent "The Invisible Web"). Ils demeurent très utiles quand les outils traditionnels échouent. Cependant, leur éparpillement rend leur utilisation souvent assez longue et difficile. Pour cette raison, certains agents de recherche se sont spécialisés par secteur d'activité (Inforia Quest, WebSeeker).

Dans le cadre du projet Viamm, nous avons choisi d'utiliser un outil de surveillance de pages off-line (e-Catch 3.0) capable d'aspirer des pages suivant différents paramétrages (sur la profondeur, les types d'information, un pourcentage de modification, ...). Un refiltrage est éventuellement effectué à l'aide d'un moteur de recherche utilisant un dictionnaire sémantique pour étendre la couverture des mots clés à rechercher. Cette extension permet par ailleurs d'augmenter la pertinence des résultats

(les pages contenant un seul mot clé sont défavorisées par rapports aux pages contenant également des mots en relation dans le thésaurus avec les mots clés constituant la requête initiale).

Pour traiter certains cas particuliers, j'ai proposé d'inclure **une rubrique "Moteurs en ligne"** au portail Viamm afin d'exploiter des moteurs sectoriels. Cette rubrique a pour rôle de repérer de nouvelles adresses qui seront surveillées dans un second temps à l'aide d'un agent pull off-line.

Afin de faciliter l'intégration de ces adresses, une **rubrique "bookmarks"** a été également ajoutée. Cette rubrique permet de mémoriser les adresses repérées et laisse à l'administrateur du portail le temps de les tester (de paramétrer notamment la périodicité souhaitées des rapatriements, la profondeur de ces rapatriements, ...).

II.1.3. Comparaison avec les offres commerciales de solution de portail pour la veille

L'étude de l'offre s'est concentrée sur trois éditeurs / intégrateurs : Mediapps, Arisem, Qwam System qui présentent des produits proches les uns des autres et parfois en collaboration. Des réunions d'information ont permis de prendre connaissance de ces produits.

II.1.3.1. Caractérisation générale de l'offre

La notion de Portail vue par les éditeurs

Un portail est un site Web à trafic et attrait élevés, offrant un large éventail de contenus, de services et de liens vers des fournisseurs de contenu. Il joue le rôle d'intermédiaire à valeur ajoutée, en sélectionnant les sources de contenu et en les réunissant sous une interface facile à explorer (et à personnaliser) pour les présenter à l'utilisateur final. Les portails incluent généralement des services tels que la messagerie, le rassemblement de communautés et la conversation en ligne.

Ce type de service inclut une grande part de « personnalisation ». Celle-ci correspond à une utilisation de profils utilisateurs constamment réajustés afin de faire correspondre un contenu ou des services aux besoins des utilisateurs. La personnalisation comprend : la détermination de l'intérêt de l'utilisateur en fonction de ses préférences ou de son comportement, la sélection d'un contenu adapté à ses préférences ou ses comportements, la présentation du contenu à l'utilisateur final sous un format cohérent et structuré.

II.1.3.2. Mediapps

NetPortal est un produit de la société Mediapps. Il se classe dans la catégorie des portails d'accès à une information personnalisée et est adaptable en intranet et extranet pour l'entreprise. L'interface est adaptée au profil de l'utilisateur et lui permet de recevoir des informations ou services personnalisés. L'offre se décline en Net Portal for Domino, Net Portal for Java, et Net Portal for IIS pour correspondre aux différentes architectures.

Son utilisation est souple et permet de construire un portail personnalisé (au niveau graphique et au niveau contenu) qui peut être utilisé comme la page d'accueil du site Internet, extranet ou intranet ou comme une extension à la manière de MyYahoo! Les informations provenant de ces différentes sources sont agrégées et accessibles via l'interface. Une diffusion de l'information par e-mail et les listes de diffusion est également possible.

L'intérêt principal de NetPortal est son offre de sources d'informations externes (grâce à des accords avec les fournisseurs et son partenariat avec QwamSystem pour le produit NetPortal) auxquelles peuvent venir s'ajouter les sources d'information internes de l'entreprise. Cette offre comprend un ensemble (environ 10000) de sources d'informations gratuites ou payantes (Les Echos, Le Monde, Yahoo, Reuters, Lexis-Nexis, LegalNews, Transfert.net, Assemblée Nationale, Ministères...). NetPortal se

positionne donc comme un intermédiaire entre les sources d'informations et les utilisateurs finaux.

Points faibles :

- le filtrage des documents est limité à l'existence ou non de mots clés dans ce qui a pu être identifié comme un titre pour le document.
- seules des alertes sont envoyées aux utilisateurs, il ne s'agit pas d'une aspiration des documents. Sous certaines conditions, les documents peuvent être aspirés. Toutefois, seul le contenu textuel est répliqué.

II.1.3.3. QwamSystem

QwamSystem est une société spécialisée dans l'édition et l'intégration de portails pour les entreprises. Qwam dispose d'une vitrine accessible par l'Internet : le portail Internet de l'information professionnelle www.qwam.com. Sur ce site, sont accessibles de manière payante ou non, 25000 sources d'informations organisées en 300 bases de données différentes. Ces sources sont interrogeables de façon homogène par l'intermédiaire d'un langage Qwam traduit en langage spécifique à la base de données interrogée. QwamSystem met à disposition de NetPortal une partie de ses contenus d'information dans les thèmes suivants : archives des brevets, information générale, sciences et techniques, études de marché, informations financières...

Qwam propose également une offre sur le marché des portails d'accès à l'information pour les entreprises mais cette fois de façon autonome, sans partenariat avec Mediapps. Le produit Qwam e-content server (voir annexe) permet d'alimenter les sites portails des entreprises en informations professionnelles internes ou externes. Pour ce produit aussi, on retrouve la possibilité de personnaliser l'interface au niveau graphique et contenu. La sélection des sources se fait parmi l'ensemble des 25000 fournisseurs d'informations disponibles par l'intermédiaire de Qwam.

Points faibles :

- Le langage d'interrogation "universel" utilisé est un compromis qui ne permet pas d'utiliser certains critères spécifiques des bases de données spécialisées.
- Le contenu doit rester sur le serveur distant et ne peut pas être rapatrié.

II.1.3.4. Arisem

Arisem est un éditeur de logiciels proposant des analyses sémantiques. Ces analyses mises en œuvre dans le cadre de la conception d'un portail peuvent permettre une augmentation de la pertinence des informations présentes dans ce portail.

En effet, les deux outils précédents n'exploitent pas la dimension sémantique de l'information. Bien que personnalisée, l'information mise à disposition par l'intermédiaire de ces outils peut parfois manquer de pertinence par rapport aux attentes des utilisateurs. En utilisant un outil d'analyse sémantique en amont d'un portail, on gagne en pertinence pour les données proposées à la consultation.

L'offre Arisem sur le segment des portails d'accès à l'information pour les entreprises se compose de quatre produits : IntraPortal4U (intranet, portail documentaire), WatchPortal4U (Internet, portail d'intelligence économique), KnowledgePortal4U (partage des connaissances sur intranet), WebPortal4U (portail Web thématique pour communautés). Un module d'analyse sémantique est aussi commercialisé à part. Il peut prendre place en amont de n'importe quel type de portail.

QwamSystem, présenté plus haut, est aussi partenaire d'Arisem sur ces produits en fournissant l'accès à ses sources d'informations professionnelles.

Points faibles :

- le coût de la solution est très important (350 KF pour l'achat du logiciel seul) ;

- la mise en place de la catégorisation des informations nécessite la définition de modèles de connaissance complexe (service proposé par les équipes de linguistes d'AriseM).
- il n'existe pas de fonctions de reformatage des résultats des recherches.

II.2. Technologies Web adoptées

Les outils choisis sont très répandus aujourd'hui. Les listes de diffusion, les sites spécialisés, les forums concernant ces technologies sont nombreux et permettent de résoudre certains problèmes.

II.2.1. Contraintes

Aux contraintes de délocalisation et de visualisation dynamique (propre à la définition même du portail Viamm) s'ajoutent deux contraintes. Afin d'encapsuler les différents outils de recherche en amont du portail Viamm, il est tout d'abord nécessaire de pouvoir visualiser dynamiquement les résultats de leurs recherches. Ensuite, des traitements doivent être effectués sur les fichiers rapatriés de façon à les mettre en conformité avec la base de données "Missile News".

II.2.2. Choix des technologies

II.2.2.1. Le langage Perl

Les signalements des documents construits par l'intermédiaire du portail Viamm, sont à exporter vers la base de données "Missiles News" dans un format particulier. Cette nécessité de reformatage a dicté le choix du langage Perl.

Perl est un langage gratuit qui reste une référence sur Internet. Rapide, puissant, c'est un langage universel que les développeurs ont adopté facilement. Il est toutefois plus complexe à mettre en œuvre que le PHP.

Ce langage est employé dans ce projet sous forme de routines permettant la compatibilité des formats entre la base de données du portail et la base de données "Missiles News".

Un outil de mise au point de traitements de fichiers accompagne le portail. Cet outil se présente sous la forme d'une interface graphique permettant de générer des scripts de traitement de données textuelles. Ces scripts réalisent par exemple des remplacements, des extractions de parties du texte ou encore des transformations à partir des expressions régulières du Perl. Ainsi, l'utilisateur peut choisir d'associer à une adresse donnée un script de reformatage personnalisé qui alimentera automatiquement certains champs d'une fiche de signalement (voir *Figure 6*).

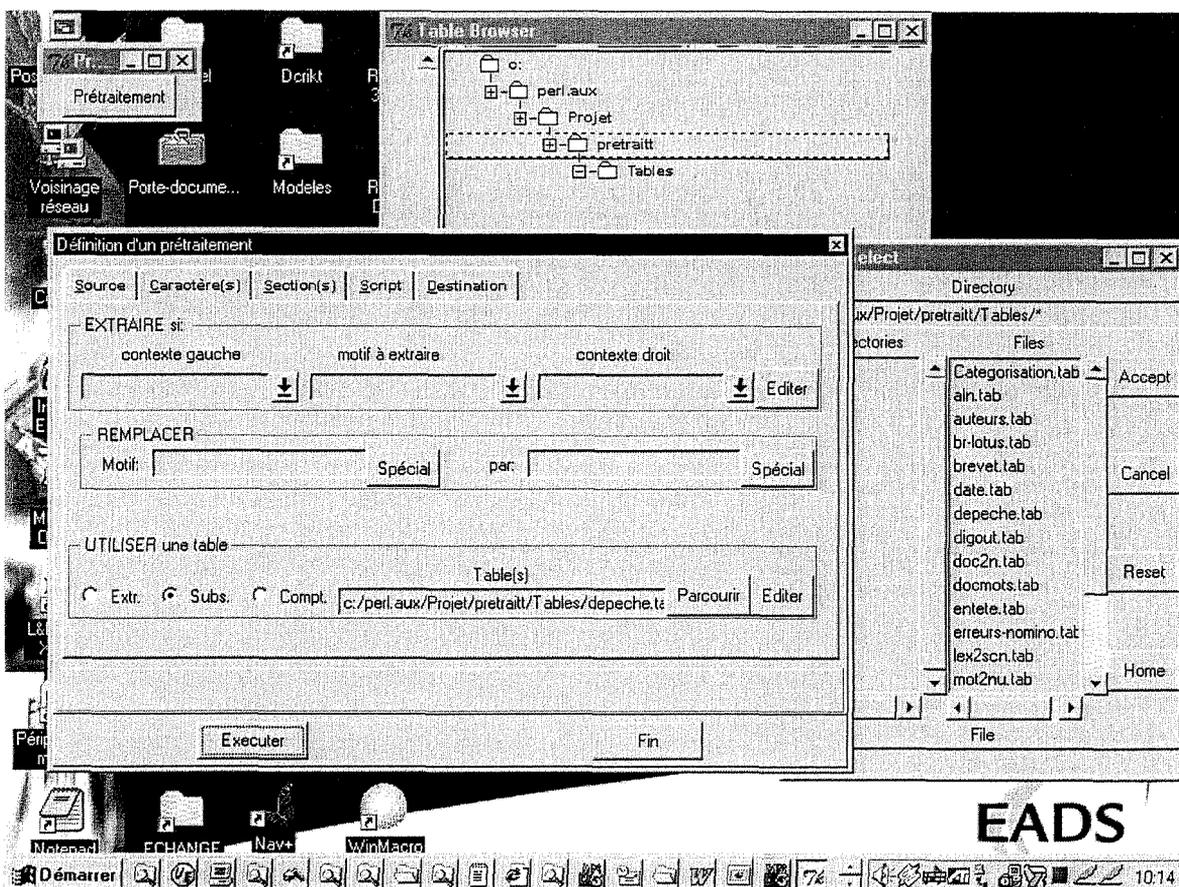


Figure 6 : Interface graphique de l'outil de traitement de fichiers en Perl

Le langage Perl utilisé est ActivePerl 5.005.03 avec les bibliothèques graphiques Tk800.015 pour Windows NT.

II.2.2.2. Le langage PHP

Le langage PHP a été choisi car la gestion des bases de données est relativement facilitée. Si, à ses débuts, il y a un ou deux ans, PHP était un langage assez confidentiel, il connaît aujourd'hui une très forte croissance [PHP2000], [DEZ2000].

La gestion dynamique des pages du portail Viamm repose entièrement sur son utilisation couplée à celle de la base de données MySQL. Un module de PHP est dédié à la gestion de la base de données MySQL, il s'agit de PHPMyAdmin. Ce module facilite la gestion de la base en proposant une interface Web accessible directement par l'intermédiaire du navigateur. Ce langage est intégré directement au langage HTML qui lui, ne peut produire que des pages statiques.

Le langage PHP permet de bénéficier d'instructions conditionnelles, de calcul... qui sont indispensables dans la mise en place de pages dynamiques. Un schéma présente le fonctionnement du langage PHP intégré au langage HTML 4.0 (voir annexe section **Erreur! Source du renvoi introuvable.**). Ce langage est gratuit [HER1999].

La version utilisée de PHP est PHP4.

II.2.2.3. Le langage JavaScript

Au sein du langage HTML, le JavaScript a aussi été utilisé pour l'amélioration de l'ergonomie, de la navigabilité du portail Viamm. JavaScript a été mis au point par Netscape. Si JavaScript peut fonctionner côté serveur, il est généralement utilisé côté client. Les instructions sont directement insérées dans les pages HTML et traitées par le navigateur. C'est un langage très largement employé sur Internet [CHA1999].

Les langages PHP et PERL sont tous deux exécutés côté serveur par opposition au JavaScript qui est exécuté côté client. Si les langages côté serveur offrent beaucoup

d'avantages, il est nécessaire de recharger la page chaque fois que celle-ci est modifiée. Avec un langage côté client, les modifications peuvent se faire sans rechargement de la page. Le portail « VIAMM » de par sa conception, mêle les scripts s'exécutant côté serveur et côté client lorsque l'un ou l'autre s'avère le plus pratique. De plus certaines tâches ne peuvent être prises en charge que par des scripts exécutés côté serveur telles que les « rollovers » qui désignent les changements d'apparence de liens, textes ou images au passage de la souris.

II.2.2.4. La base de données MySQL

La base de donnée MySQL est la base dans laquelle seront stockées toutes les données transitant par le portail Viamm avant de rejoindre la base "Missile News". C'est l'une des base de données compatibles avec le langage PHP qui possède des fonctions spécifiques à son interrogation. L'introduction de requêtes SQL interrogeant cette base en est facilitée. Elle permet la gestion des droits alloués aux utilisateurs ou administrateurs. Ce système de gestion de bases de données compatible avec le Web via PHP est gratuit.

La version utilisée est MySQL 3.22.29-Win.

II.2.2.5. Le serveur Apache

Le serveur Apache fédère l'ensemble des technologies précédemment choisies moyennant quelques adaptations au niveau des fichiers de configuration. Il permet les échanges client serveur rendus nécessaires par l'architecture du portail Viamm. Ce serveur permet d'étendre, si besoin est, l'accès du portail Viamm à des postes connectés au réseau local du centre de documentation. En effet, ce serveur est accessible sur le poste de travail depuis un navigateur par l'intermédiaire d'un localhost.

La version utilisée est Apache 1.3.9. pour Windows NT. Il est téléchargeable gratuitement depuis le Web.

III. Conduite du projet

La conduite des projets au Centre Commun de Recherche est soumise à des contraintes qualité dérivée des normes ISO 9001. Chaque projet dans lequel apparaît une tâche de réalisation logiciel repose sur un plan de management qui lui-même s'appuie sur un dossier d'analyse (Rapport sur les technologies utilisées, Modélisation des flux d'information) et un dossier de validation (Plan de test et résultats)¹.

Le rôle du plan de management est d'évaluer les risques afin de les prévenir au mieux. Cette formulation quelque peu négative peut être reformulée en "management des facteurs de succès". Chaque étape du processus de réalisation qui m'a été confié a fait l'objet d'une validation puis d'une présentation au client. Ces étapes étaient :

- l'étude des technologies à utiliser pour le portail Viamm ;
- la conception de l'ergonomie du portail Viamm (décomposition en rubriques et enchaînement des pages) ;
- le développement des composants ;
- le test des composants ;
- l'intégration des composants ;
- la finalisation (test sur des données réelles, corrections des scripts)

¹ Se reporter à la section V.1 pour plus d'informations sur les scénarios d'utilisation ayant servi pour le plan de test

Ces étapes s'inscrivent dans un découpage plus large inspirés de plusieurs méthodes telle la méthode CALS (*Computer Acquisition and Life Cycle Support*), issue du département de la Défense américain. Elles sont illustrées *Figure 7*.

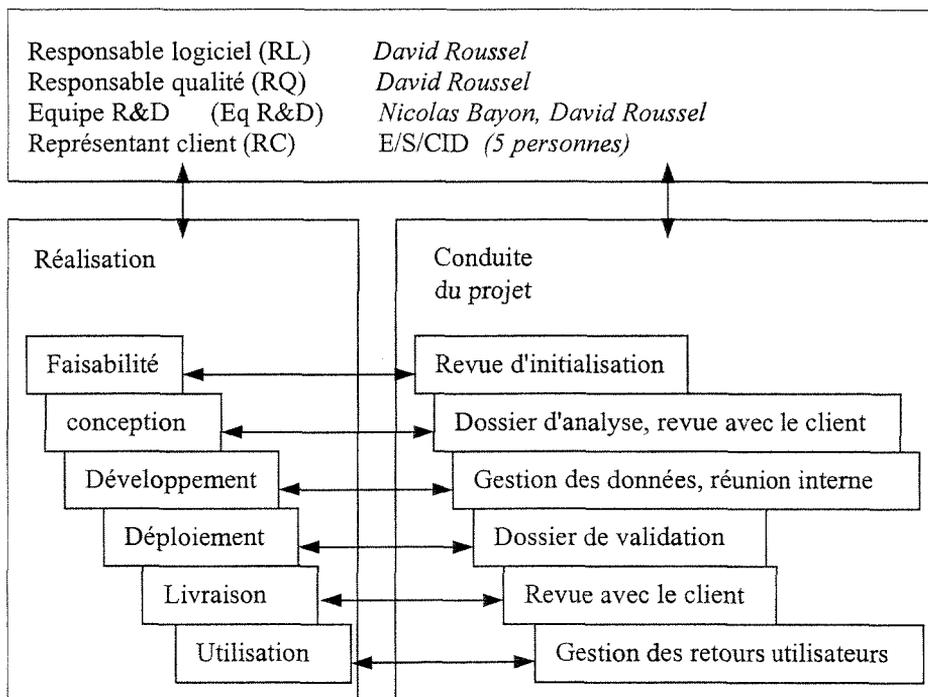


Figure 7 : Méthode d'organisation suivie pour le projet VIAMM

La revue d'initialisation a pour rôle de mettre en place les objectifs du démonstrateur. Cette revue est décrite dans un Compte rendu d'ouverture qui était préalable au début de mon stage.

Les réunions internes sont provoquées à l'initiative du responsable logiciel. Les participants sont le responsable logiciel, le responsable qualité, l'équipe de R&D et si besoin le chef de service.

Le but de ces revues est de prévenir les non-conformité : vérifier la cohérence de l'architecture et la conception des fonctions du démonstrateur ; vérifier les résultats des tests de validation.

Les **revues de validation avec le client** sont planifiées dans le programme d'étude. Elles sont précédées d'une revue de validation interne. Afin d'éviter toute ambiguïté, un compte rendu en est donné au client, auquel il peut demander des corrections. Si des outils payant supplémentaires sont à prévoir, ceux-ci sont soumis au client lors des réunions d'avancement.

Lors de cette revue, une présentation de l'état d'avancement est faite au client, sous la forme d'une présentation de maquette logicielle dans le cas de mon projet de stage.

Une maquette initiale simulée sous forme de pages Web non dynamique a été approuvée lors d'une première revue antérieure au début de mon stage. J'ai ensuite participé à trois revues de validation avec le client, montrant chaque fois l'avancement du portail Viamm et les fonctionnalités annexes ajoutées au cours de la réalisation.

La dernière revue (11 Septembre) a consisté en l'installation de la version définitive du portail, approuvée par le client.

La gestion des retours utilisateurs a été intégrée au portail Viamm sous la forme d'un système de recueil d'avis et de votes. La réalisation de ce système n'a pas fait parti de mon stage.

La gestion des données concerne l'archivage quotidiens des différents éléments du projet identifiés par un numéro de version le cas échéant. Les éléments archivés sont les suivants :

- Démonstrateur (sources, référence ressources de développement, ...)
- Dossier d'analyse
- Dossier de validation.

Evidemment, l'enchaînement des différentes étapes ne s'est pas faite linéairement. Elle a suivi au mieux un cycle de développement résumé dans la *Figure 8*.

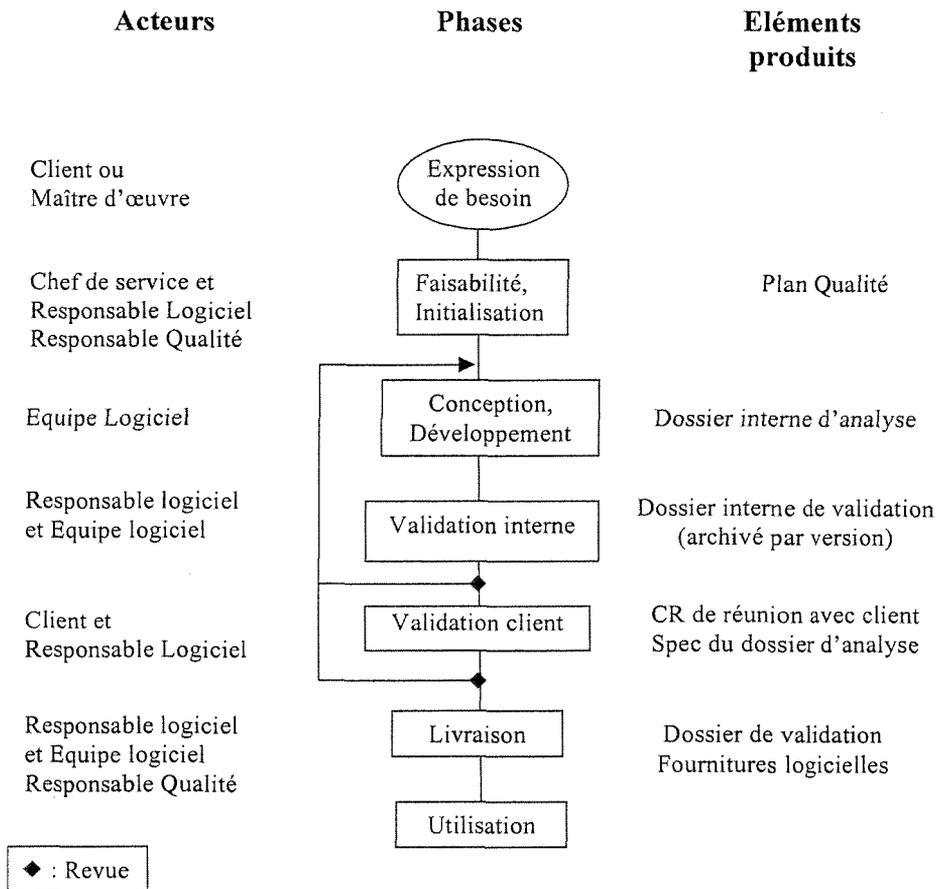


Figure 8 : Cycle de développement du portail VIAMM

IV. Réalisation du portail VIAMM : de l'illustration HTML à la maquette VIAMM V4

La réalisation du portail Viamm a suivi une évolution progressive vers une maquette de plus en plus élaborée prenant en compte les remarques formulées par les utilisateurs finaux du portail lors des réunions de travail.

IV.1. Une architecture modulaire

Il est prévu de pouvoir changer certains outils utilisés en complément de Viamm sans toucher au fonctionnement du portail.

De par sa conception, le portail Viamm peut s'adapter à différents outils amonts "sur étagère" afin de suivre l'évolution du marché². En effet, il suffit que ces outils déposent leurs résultats dans de répertoires prédéfinis pour que le portail prenne en compte ces résultats. La base de données MySQL utilisée par le portail permet par ailleurs de gérer des relations entre des adresses et des traitements particuliers.

IV.1.1. Organisation du disque dur du poste

Le stockage des documents aspirés s'effectue dans des dossiers (sur le disque dur du poste d'utilisation du portail) représentant la hiérarchie des catégories et sous-catégories de la base Missiles News (voir *Figure 9*). Les documents réapparaissent donc dans le portail VIAMM, classés suivant cette catégorisation. Cette dernière peut évoluer dans le futur mais les adaptations pour l'administrateur du portail seront minimales puisqu'il lui suffira d'adapter la hiérarchie des dossiers de stockage à ces changements en même temps que les paramètres de rapatriement des outils amonts.

² Récemment, l'outil StrategicFinder propose par exemple des accès à des bouquets d'informations spécialisées qui en font un complément intéressant de Copernic Pro 2000.

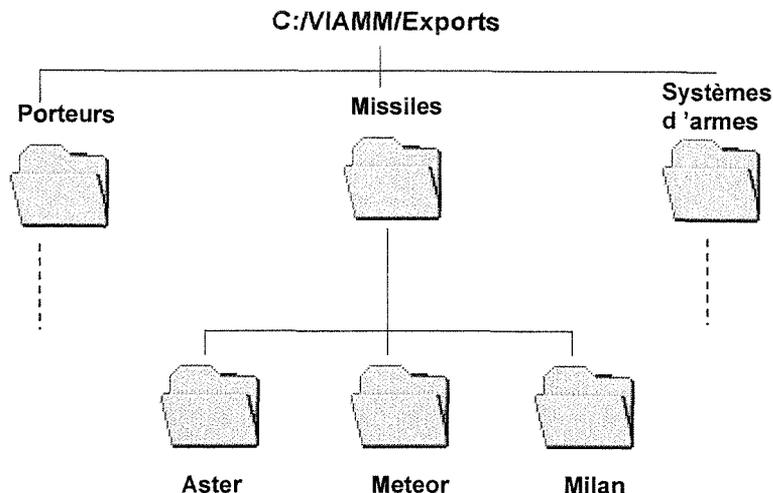


Figure 9 : Arborescence des dossiers de stockage des documents rapatriés vers le portail Viamm

Sous C:/VIAMM, il est également convenu de stocker les informations nécessaires à la collecte d'information comme suit :

- 1 répertoire *Références* : contenant des listes d'adresses sous forme de fichiers
- 1 répertoire *Données* : contenant des données rapatriées qui n'ont pas besoin d'autres traitements
- 1 répertoire *Rapatriements* : stockage intermédiaire pour les données rapatriées
- 1 répertoire *Export* : données à exporter dans Viamm. Ce répertoire est alimenté après reformatage des documents rapatriés.

IV.1.2. La structure de la base de données

La base de données liée au portail VIAMM est composée de trois tables (voir *Figure 10*):

- Bookmarks (6 champs)
-
- Traitements (5 champs)
-
- Today (11 champs)

La table "Bookmarks" rassemble les informations spécifiques aux sites sélectionnés et accessibles par le menu déroulant du cadre de gauche du portail. Ces sites peuvent être rangés dans une catégorie "En Test" qui permet à l'administrateur des outils en amont de Viamm de déterminer la pertinence de ces sites ainsi que leur périodicité de mise à jour (rafraîchissement des informations) avant de les inclure définitivement comme points de départ de nouvelles recherches par les outils amont.

La table "Today" rassemble les informations issues chaque jour des fichiers d'export des outils utilisés que sont Copernic, e-Catch et DigOut4U.

La table "Traitements" est liée aux deux précédentes tables. Elle rassemble les informations communes aux deux autres types d'enregistrements.

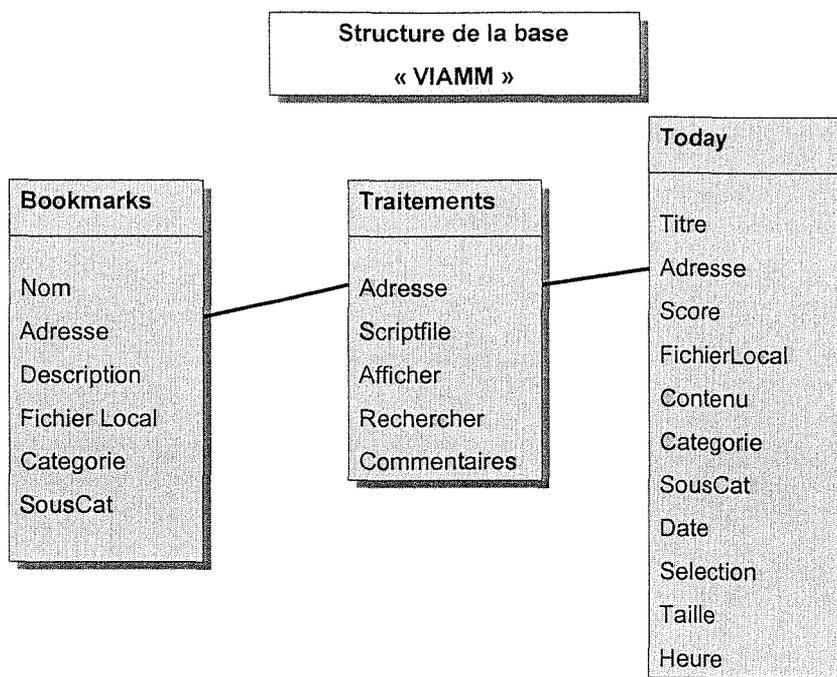


Figure 10 : Structure de la base VIAMM présente dans MySQL

La clef primaire dans chacune des tables est le champ "Adresse" qui apparaît comme champ ayant le contenu le plus solide du point de vue de la reconnaissance des documents. Le contenu de chaque fiche descriptive de chacun des documents présents dans la base est issu de cette structure.

On considère que dans cette base, un même document peut apparaître plusieurs fois dans des catégories ou même sous-catégories différentes. Ces documents ayant une adresse identique, on ajoute une seconde clé qui est la localisation des fichiers rapatriés sur le disque dur. Cette localisation est avec l'adresse (url) l'élément discriminant pour différencier les documents.

IV.2. Une interface simple

IV.2.1. Structure générale du démonstrateur

La définition de deux cadres (gauche et droite) permet à l'utilisateur un accès rapide à toutes les pages.

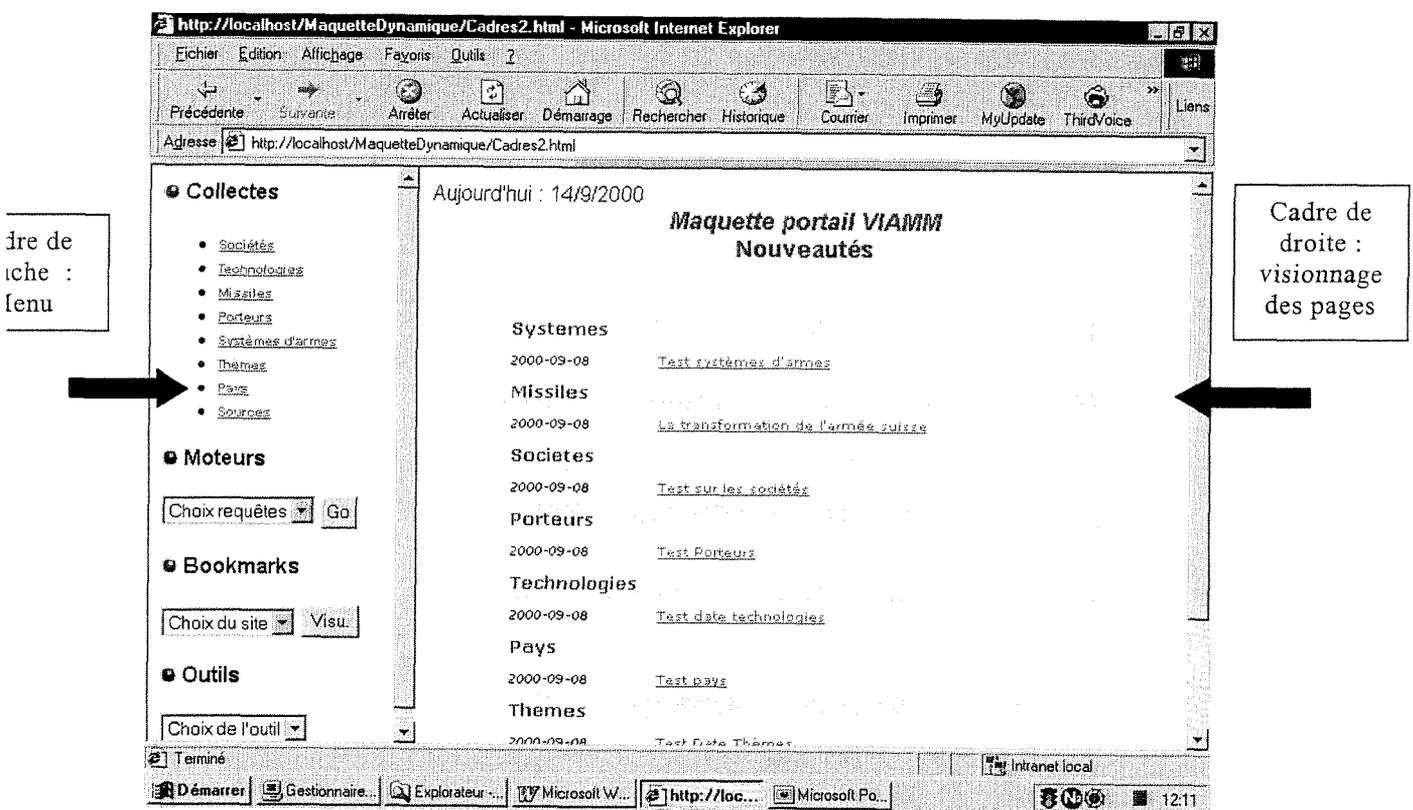


Figure 11 : Copie d'écran de la page d'accueil du portail VIAMM

Le menu, qui apparaît dans le cadre de gauche, permet l'accès de l'utilisateur à 4 rubriques :

- les collectes (résultats du rapatriement des pages Web par les outils de veille amont, traitement de ces résultats, signalement des données concernant ces pages...)

- les moteurs (accès à des recherches plus ponctuelles)
- les bookmarks (accès direct à des sites prédéterminés)
- les outils (accès à une documentation sur les outils de recherche)

La page d'accueil du portail Viamm (voir *Figure 11*), présente aux utilisateurs les références des documents les plus récents parmi ceux rapatriés lors de la dernière recherche. Ils apparaissent triés par catégorie et leur visualisation à partir du disque local est disponible par un clic sur le titre.

IV.2.2. Ergonomie des rubriques

La navigation au sein du portail a été pensée pour rester la plus instinctive possible. L'enchaînement des pages concernant chacune des grandes fonctionnalités du portail est présenté dans les figures suivantes.

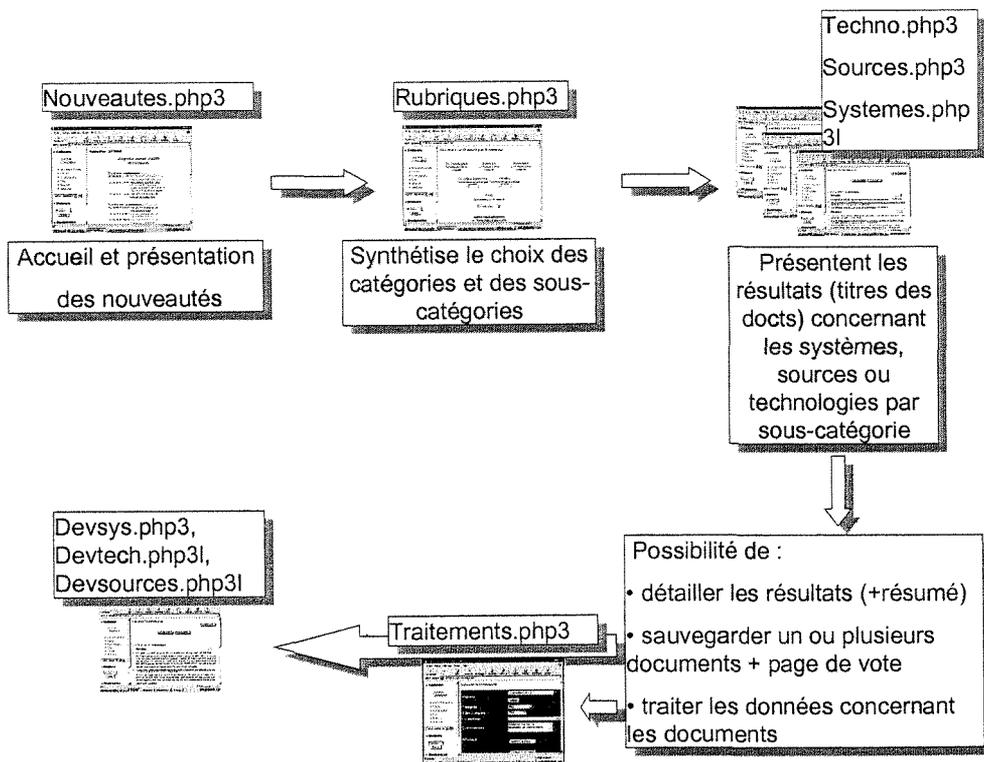


Figure 12 : Enchaînement des pages du portail à partir du menu "Collecte"

IV.2.2.1. Page de présentation des collectes

Le menu collecte donne accès au signalement et au traitement des documents rapatriés par les outils de veille. L'utilisateur peut choisir d'y accéder par la page de présentation des catégories.

• Pages de présentation des catégories

Cette page présente les sous-catégories de la base "Missiles News" sous forme de menus déroulants (qui sont des liens directs vers ces sous-catégories) permettant une visualisation rapide des signalements (voir *Figure 13*). Il existe 8 catégories et environ 200 sous-catégories.

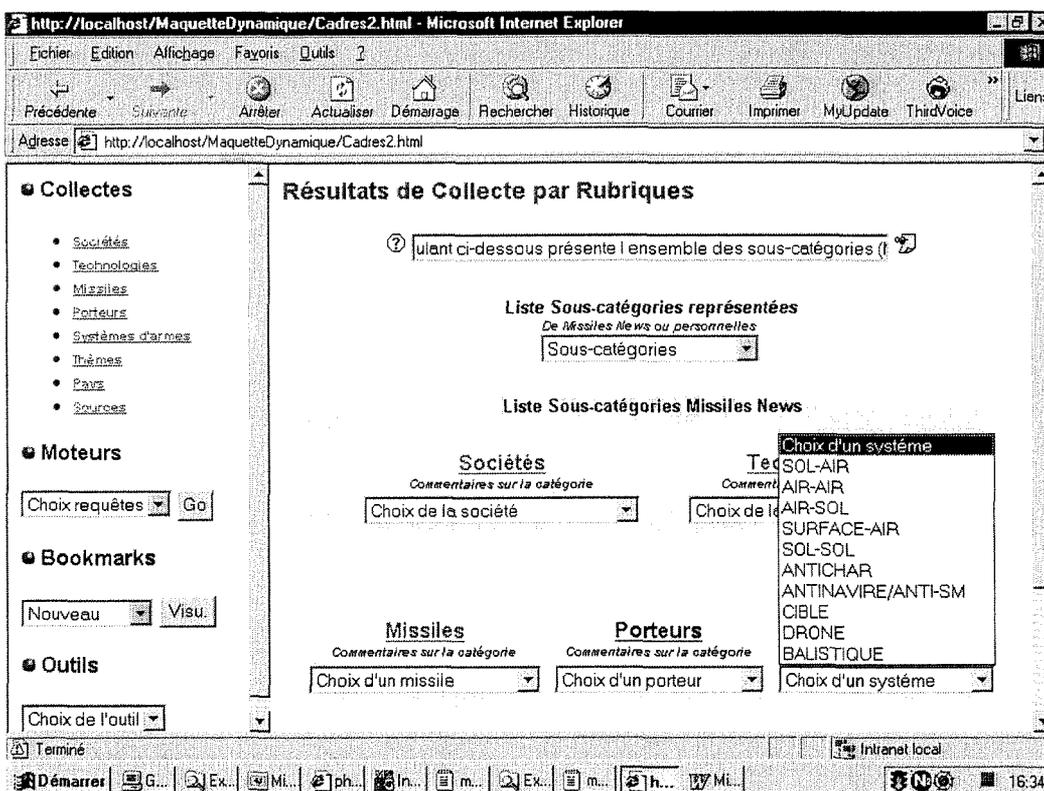


Figure 13 : Page "Résultats des collectes par rubriques" du portail VIAMM

Une liste des sous-catégories représentées dans la base est également disponible.

• *Vue des collectes par catégorie*

La présentation des données concernant les résultats de recherche se fait par l'intermédiaire des pages de "Vue" des différentes catégories (voir *Figure 14*).

La présentation de la page est identique pour chacune des 8 catégories. On y retrouve la sous-catégorie du document, son titre, sa date de création ou de mise à jour, son résumé est accessible en mode "Détailler", sa taille matérialisée par des icônes (1 ou plusieurs pages), sa pertinence par rapport à la recherche engagée par l'outil qui l'a rapatrié.

Fonctionnalités particulières aux pages "Vue par catégories" :

- Visualisation du document stocké sur le disque local en cliquant sur l'icône de taille (le document s'ouvre dans une nouvelle fenêtre) ;
- Accès au résumé par le bouton "Détailler" et inversement avec le bouton "Condenser"
- Visualisation de la page de signalement du document en cliquant sur le titre ;
- Sélection / Désélection de tous les documents par le bouton "Tous / Aucun" .

C'est sur cette page que l'utilisateur doit choisir les documents qui lui semblent intéressants pour une intégration de leur signalement à la base "Missiles News". Toutes les fonctionnalités ci-dessus lui permettent d'évaluer le document et les données qui lui correspondent.

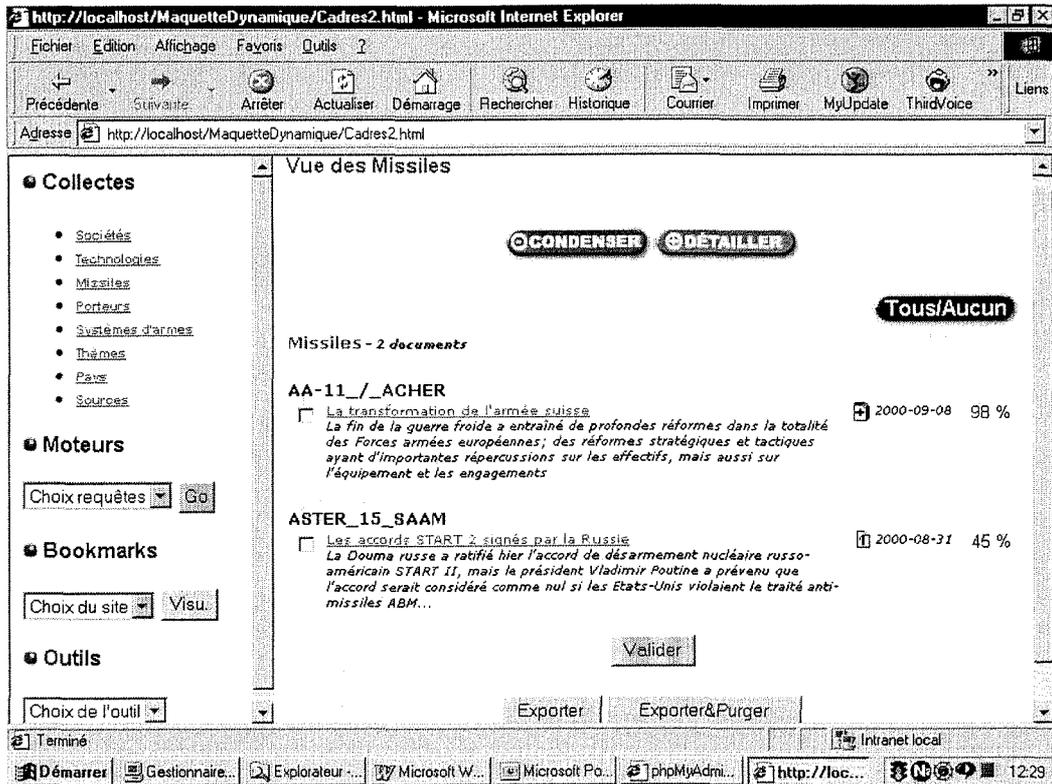


Figure 14 : Page de vue des résultats de recherche par catégories (exemple Vue des missiles)

• Page de signalement

Des modifications peuvent être apportées aux données décrivant les documents. Elles peuvent être effectuées sur la page de signalement où l'utilisateur peut intervenir sur tous les champs présents sur cette page (résumé, date, adresse, catégorie, sous-catégorie)(voir *Figure 15*). Un bouton "Enregistrer" en base de page permet de mettre à jour la base de données.

Le bouton "Visualiser" permet d'accéder à la page "on-line" à partir de l'url présente dans le champ "Adresse".

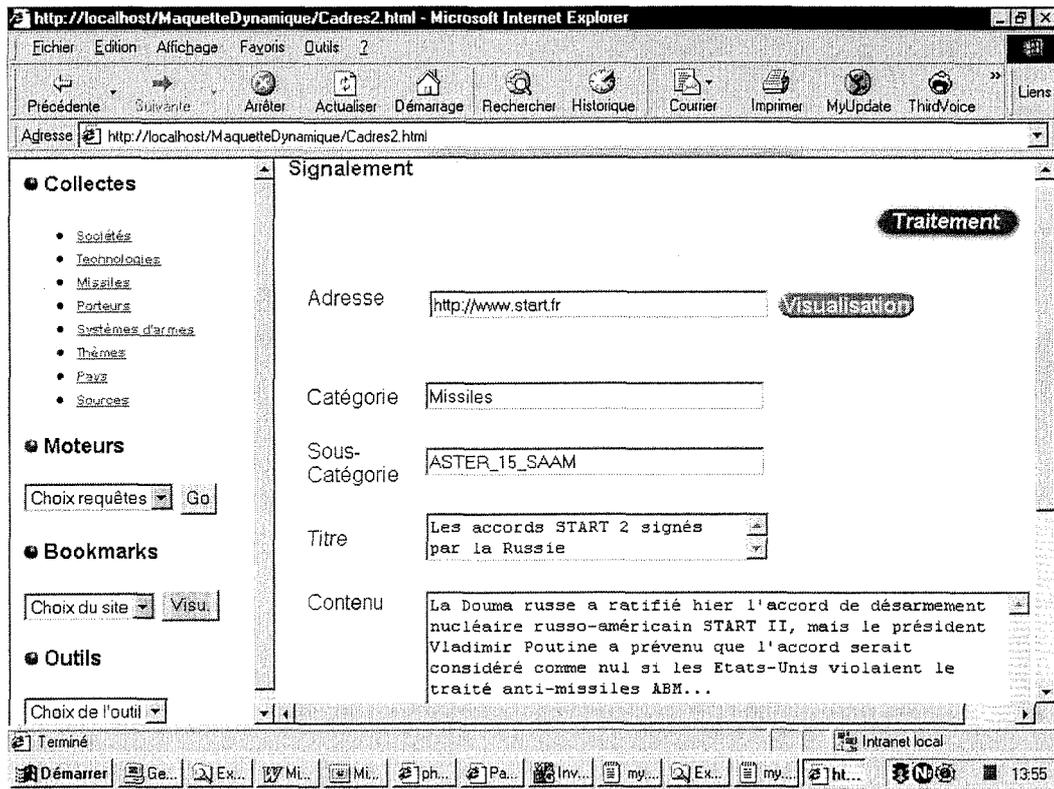


Figure 15 : Page de signalement d'un document classé dans la catégorie missiles

Le bouton "Traitement" permet quant à lui, l'accès au choix des traitements à appliquer aux documents rapatriés.

- *La page de traitement*

Elle rappelle les données relatives à un document avec les champs Titre, Adresse, Catégorie, Sous-Catégorie, la localisation du document sur le disque...(voir **Figure 16**) Toutes ces données sont modifiables directement dans le champ texte. Mais il existe un plus du côté du choix des catégories et sous-catégories. Ces champs sont des menus déroulants : la sélection d'une catégorie entraîne la modification du menu déroulant des sous-catégories de façon dynamique. Ainsi, si l'utilisateur choisit la catégorie "Missiles", la liste des sous-catégories "Missiles" apparaîtra dans le menu déroulant correspondant. Il faut noter que l'utilisateur peut ajouter une nouvelle sous-

catégorie en demandant "autre" (une boîte de dialogue apparaît) dans le menu déroulant des sous-catégories. Dans la base de données, le nom des catégories est contrôlé mais pas celui des sous-catégories.

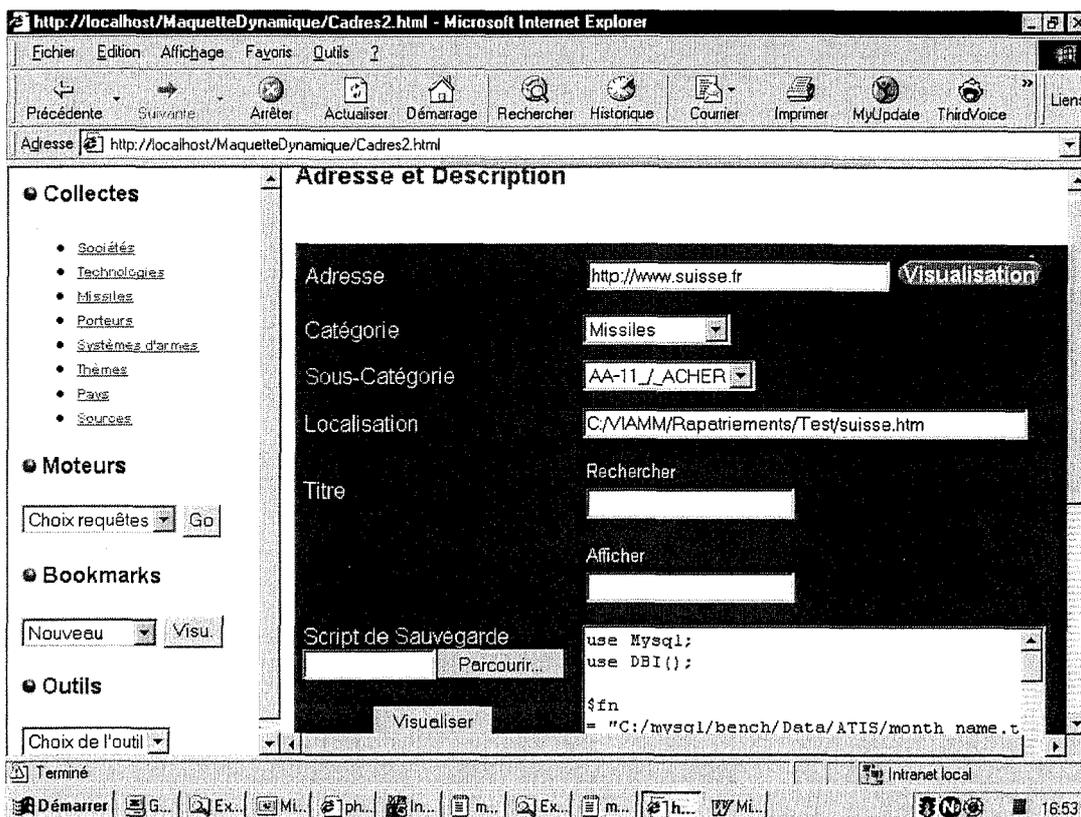


Figure 16 : Page de spécification du traitement à effectuer sur les documents répertoriés dans le portail VIAMM

Cette page permet également d'ajouter de nouvelles informations sur les documents. Elle permet notamment de choisir un type de traitement à effectuer sur les documents par un bouton "Parcourir" par lequel l'utilisateur accèdera à une boîte de dialogue lui permettant de choisir un fichier de traitement en fonction de la structure du document auquel il a à faire. Le script du traitement à effectuer s'affiche dans le champ texte qui le juxte pour vérification. Les champs "Afficher" et "Rechercher" sont relatifs au titre du document et spécifient des actions à effectuer sur le titre pendant le traitement.

La visualisation du document "on-line" est possible via le bouton "Visualisation".

Des commentaires personnels peuvent être ajoutés dans un champ réservé à cet effet.

Le bouton "Enregistrer" permet de mettre la base de données à jour par rapport aux modifications effectuées.

- *Sélection de documents*

Après avoir pris connaissance et modifié (si besoin) les données concernant les documents rapatriés, l'utilisateur doit effectuer une sélection de ces documents. Pour cela, il revient à la page de présentation des documents par catégories (voir *Figure 14*).

En bas de page, trois boutons sont à sa disposition :

- Le bouton "**Valider**" : il permet, après avoir sélectionné un document à l'aide de la case à cocher, de préparer les données à l'export. La validation peut s'effectuer en plusieurs fois puisque les documents non sélectionnés réapparaissent sur la page. Une autre phase de choix de traitement et de modifications de données peut commencer.
- Le bouton "**Exporter**" : il permet de générer un fichier d'export depuis la base de données MySQL comprenant les données des documents sélectionnés (données modifiées ou non et type de traitement à effectuer). Cette phase d'export peut s'effectuer depuis n'importe quelle page de présentation par catégorie pour l'ensemble des documents sélectionnés. Ce type d'export peut s'effectuer en plusieurs fois : chaque nouveau fichier d'export prenant en compte les données déjà exportées. Les données sont conservées dans la base mais n'apparaissent plus au niveau du portail (évite les doublons).
- Le bouton "**Exporter&Purger**" : cette procédure d'export doit être utilisée avec précaution. En effet, un clic sur ce bouton génère un nouveau fichier d'export comme pour le bouton "Exporter" mais effectue aussi une purge de la base. Les

documents déjà exportés sont éliminés de la base alors que jusque là ils y étaient conservés même s'ils n'apparaissent plus au niveau du portail. En utilisant le bouton "Exporter&Purger", les documents sont éliminés catégorie par catégorie selon la page de présentation par catégorie d'où provient ce clic.

IV.2.2.2. L'accès aux moteurs de recherche en ligne

Pour intégrer de manière encore plus globale la collecte d'informations sur Internet, l'utilisateur du portail Viamm a la possibilité de rechercher de nouveaux sites en interrogeant différents moteurs de recherche (voir *Figure 17*).

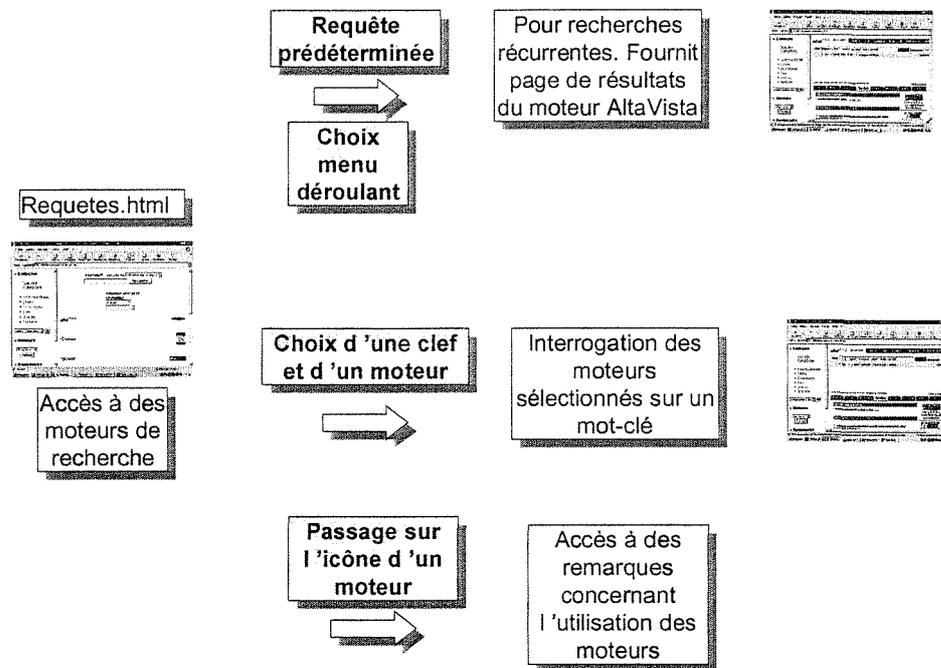


Figure 17 : Fonctionnalités principales de la page "Moteurs de recherche en ligne"

Cette recherche peut s'effectuer tout d'abord de manière directe en utilisant les champs de saisie des requêtes de la page "Moteurs de recherche en ligne" (voir *Figure 18*). Sur cette page sont disponibles un champ de requête spécifique à AltaVista où

l'utilisateur peut taper des requêtes contenant plusieurs mots clés et un champ de saisie pour un seul mot clé avec un choix de moteur de recherche par menu déroulant.

Pour surveiller l'apparition de nouveaux sites dans des domaines précis, un menu déroulant permet de lancer une recherche sur AltaVista à partir d'une requête préétablie.

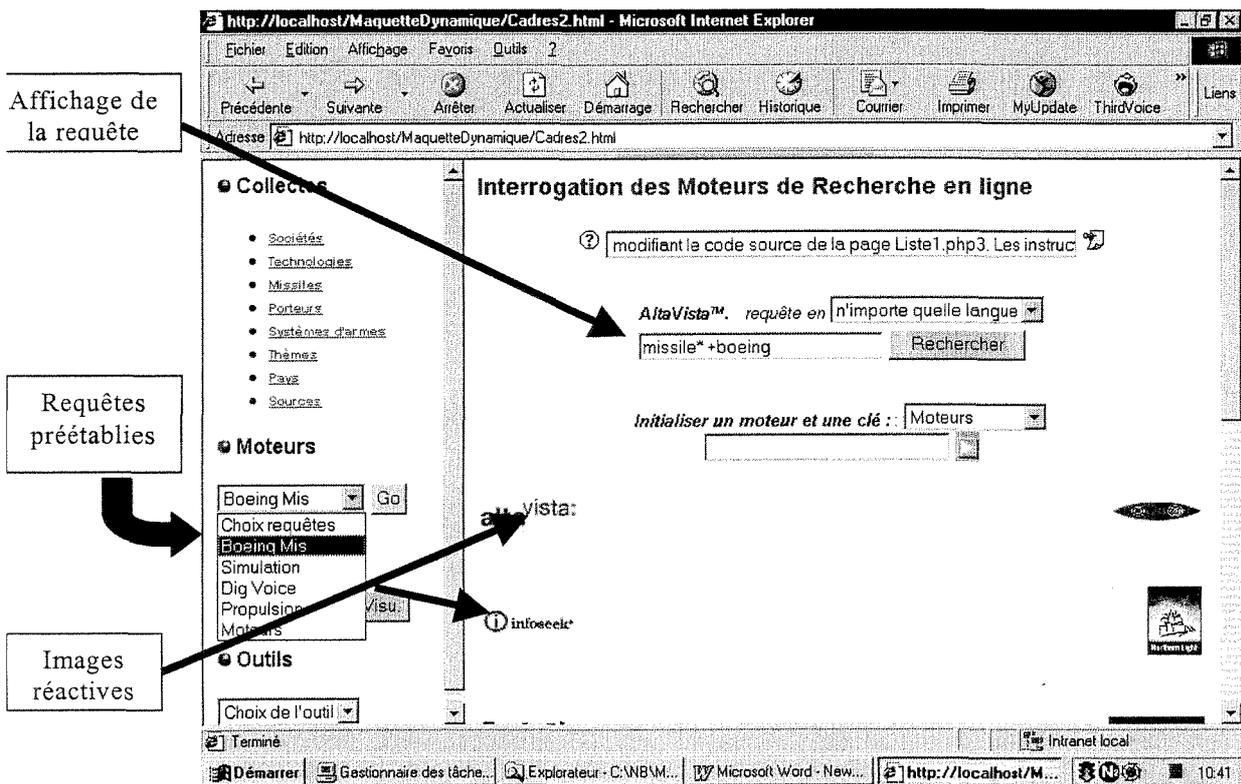


Figure 18 : Page "Moteurs de recherche en ligne" du portail VIAMM

La page de résultats concernant la requête lancée s'affiche dans le cadre de droite.

Des informations concernant l'utilisation des moteurs de recherche (syntaxe des requêtes, spécialisation dans un domaine...) est disponible sur cette page au passage de la souris sur les logos (images réactives) correspondant aux moteurs.

Dans la liste des moteurs disponible à l'interrogation, certains sont très spécialisés et peuvent rapporter des sites intéressants à tester comme point de départ de nouvelles recherches pour les outils amont (voir page de résultats du moteur SearchMil (moteur spécialisé dans le référencement des sites militaires) *Figure 19*).

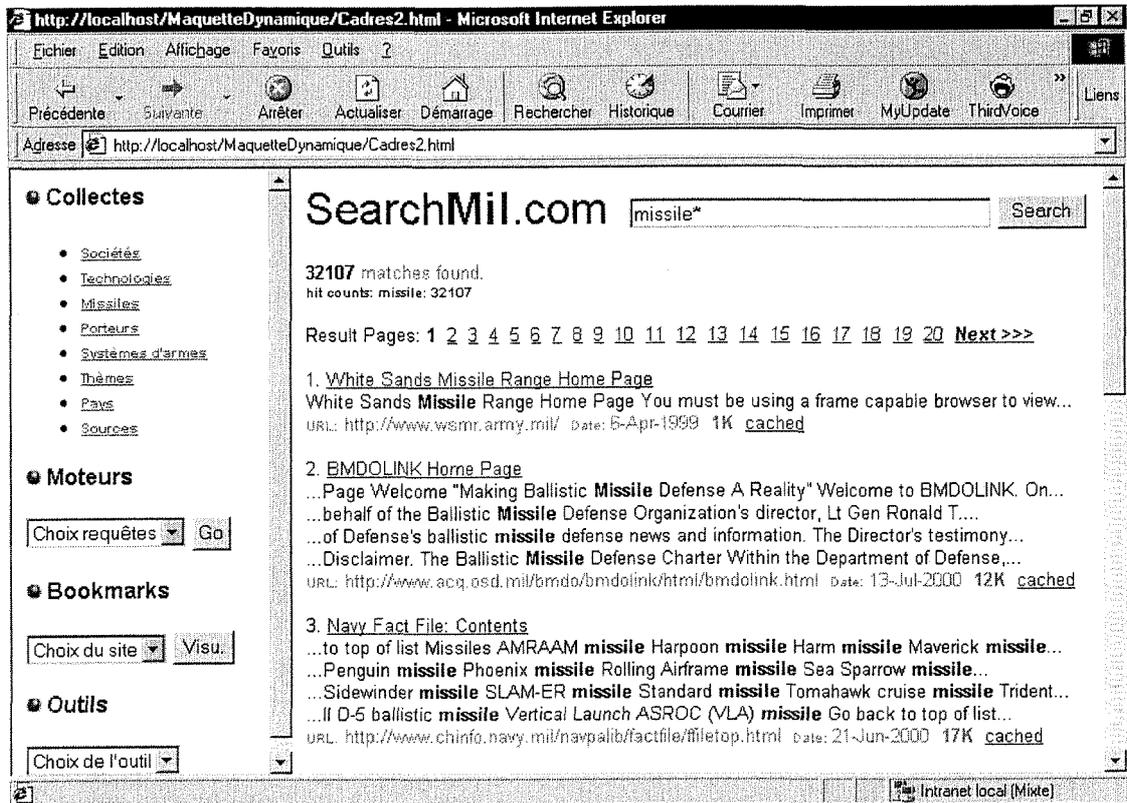


Figure 19 : Résultats de recherche après une requête sur SearchMil

Des tests doivent être effectués sur ces nouveaux points de départ potentiels de recherches. Avant de les intégrer aux outils amonts, cette évaluation peut s'effectuer à partir de la rubrique "Bookmarks".

IV.2.2.3. Les bookmarks

Cette rubrique permet de référencer des sites intéressants pour les utilisateurs. Ces sites peuvent se situer à la périphérie des sujets de recherche de la base "Missiles News" : il ne peuvent pas être classés sous forme de catégories et sous-catégories et faire partie intégrante de la base mais doivent tout de même être consultés régulièrement par les documentalistes.

Ces sites peuvent aussi être des points de départ potentiels pour de nouvelles recherches des outils amonts. Cette rubrique permet de les évaluer pendant une période donnée, de juger de leur pertinence par rapport aux sujets recherchés et déterminer leur

périodicité de rafraîchissement ou de mise à jour des informations (ces résultats permettant ensuite de paramétrer les outils amonts).

• *Description du Bookmark*

La visualisation de ces sites se fait grâce à un menu déroulant. On accède à une description rapide du site ou de la page (adresse, nom, description, voir *Figure 20*).

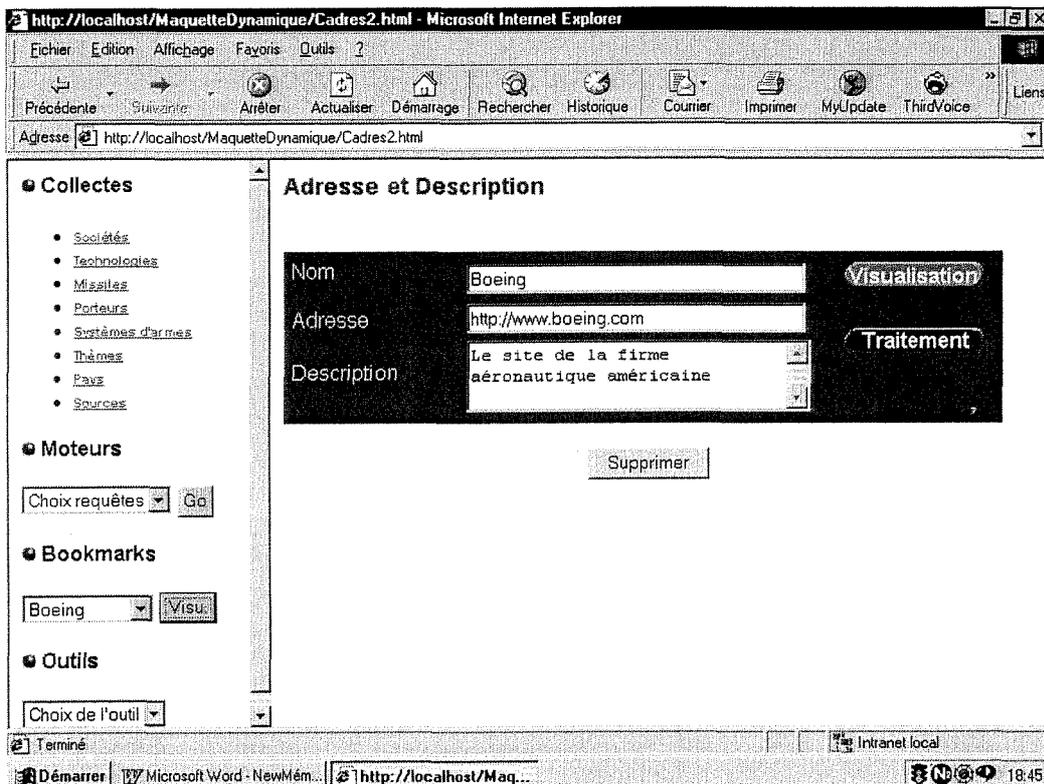


Figure 20 : Page de description d'un site référencé dans la rubrique "Bookmarks"

Fonctionnalités de cette page :

- Le bouton "Visualisation" : il permet d'accéder au site "on-line"
- Le bouton "Traitement" : il permet d'accéder à la page de traitement identique à celle des documents de la rubrique "Collecte" (voir *Figure 16*). Il est possible d'affecter à ces sites une catégorie parmi les 8 présentes dans la base de données. Une particularité au niveau des bookmarks est la catégorie "EnTest" qui correspond à un site à tester avant de l'intégrer comme point de départ de recherches pour les outils de veille

- Le bouton "Supprimer" : il permet d'effacer de la base de données la référence à ce site. Il n'apparaîtra plus dans le menu déroulant du cadre de gauche.

• *Nouveau Bookmark :*

En choisissant l'option "Nouveau" dans le menu déroulant "Bookmarks", l'utilisateur peut entrer de nouvelles références. Il accède à la page "Nouveau Bookmark" et renseigne les champs adresse, nom et description.

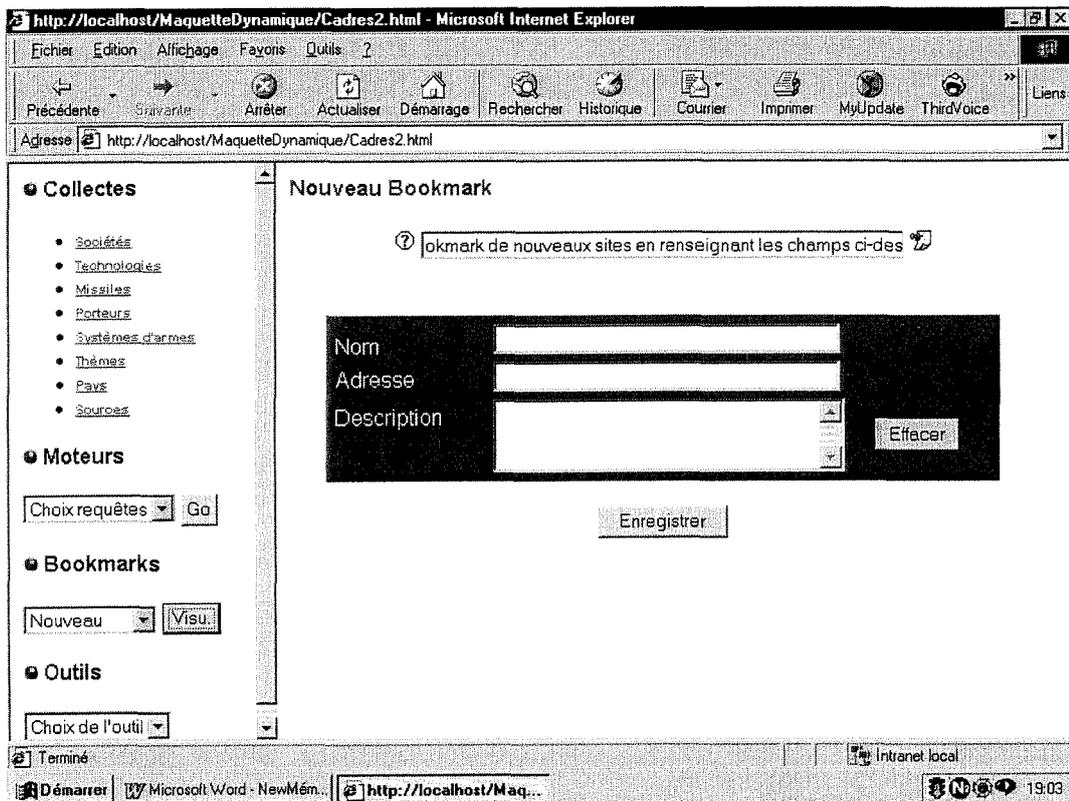


Figure 21 : Page "Nouveau Bookmark"

Le bouton "Enregistrer" permet de mettre à jour la base de données.
 Le bouton "Effacer" permet de remettre à zéro tous les champs.

IV.2.2.4. La rubrique "Outils"

Le menu général de la rubrique "Outils" permet de visualiser en ligne le tableau comparatif des outils de veille. Des liens hypertextes dans ce tableau permettent un renvoi à une documentation plus générale sur leur fonctionnement (voir *Figure 26*).

Un menu déroulant automatise l'accès à chacun de ces documents [ROU&BAY 2000a], [ROU&BAY 2000b], [ROU&BAY 2000c].

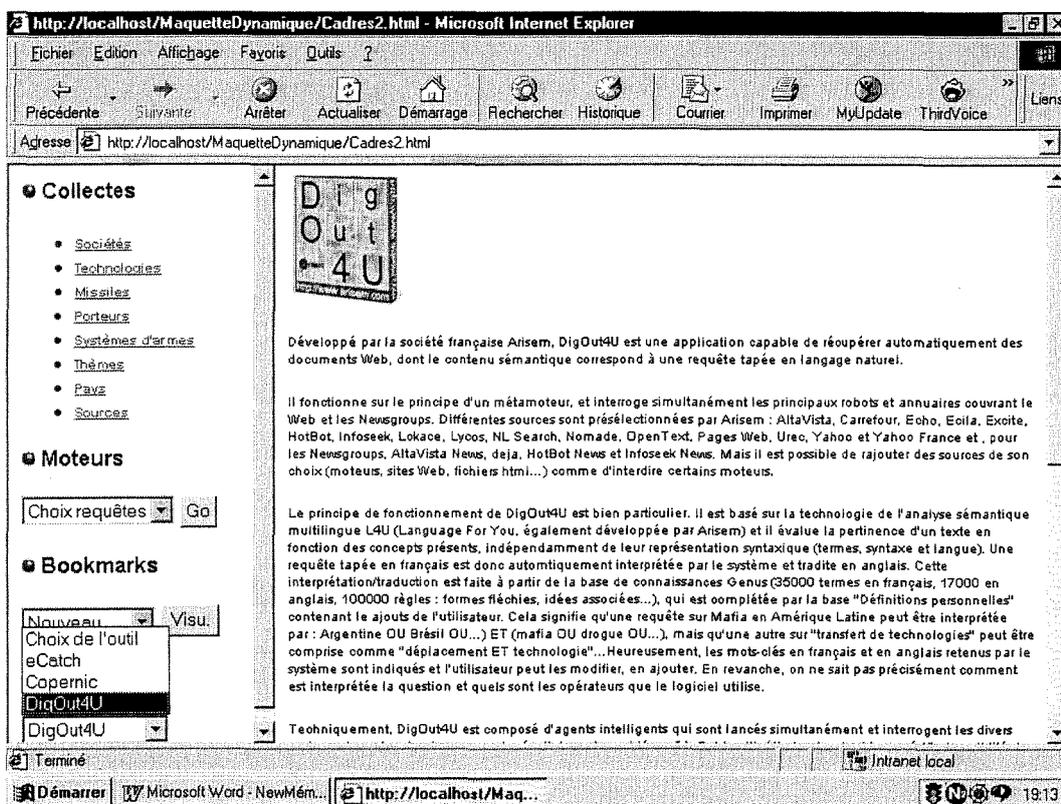


Figure 22: Une documentation en ligne est disponible dans la rubrique "Outils"

IV.3. Une gestion souple des flux de données

La plus grande partie des rapatriements d'informations pour le portail Viamm se fait de façon automatisée grâce à la mise au point de macros. Ces dernières prennent en charge les tâches les plus répétitives à réaliser par les utilisateurs. Sans ces macros, les tâches de lancement de nouvelles recherches devraient être réalisées selon une périodicité propre à chacune de ces recherches qui peut aller de « quotidienne » à « hebdomadaire », voire « mensuelle ». Le logiciel *WinMacro*, couplé au « calendrier des tâches » de Windows NT, nous a permis d'automatiser toutes ces tâches.

Par exemple, en amont du portail « VIAMM », le lancement des recherches d'e-Catch et leur refiltrage par DigOUT4U se fait par la macro « ecatchdigout ».

La macro « ecatchdigout » automatise l'activation des agents de recherches d'e-Catch (le choix des adresses est prédéterminé). Les données rapatriées sont exportées sous un répertoire du dossier *Données* (portant de préférence le nom de l'agent).

Une analyse par DigOut4U est alors lancée sur ces documents afin d'extraire les documents les plus pertinents pour une catégorie donnée. DigOut4U fournit un rapport de recherche recensant l'ensemble des documents analysés en rappelant le titre, le résumé, la date de mise à jour du document.

Ce rapport est stocké dans le répertoire correspondant à la catégorie voire la sous-catégorie des documents recherchés. Il est la base des données importées dans la table « Today » du portail « VIAMM ».

Ces données ne sont pas les seules : elles sont mêlées aux données concernant les catégories, sous-catégories, taille qui sont, elles ajoutées au moment de la modification de ce fichier de rapport de recherche à l'aide d'une routine Perl qui transforme le rapport en un fichier csv importable depuis la base de données MySQL (voir **Figure 23**).

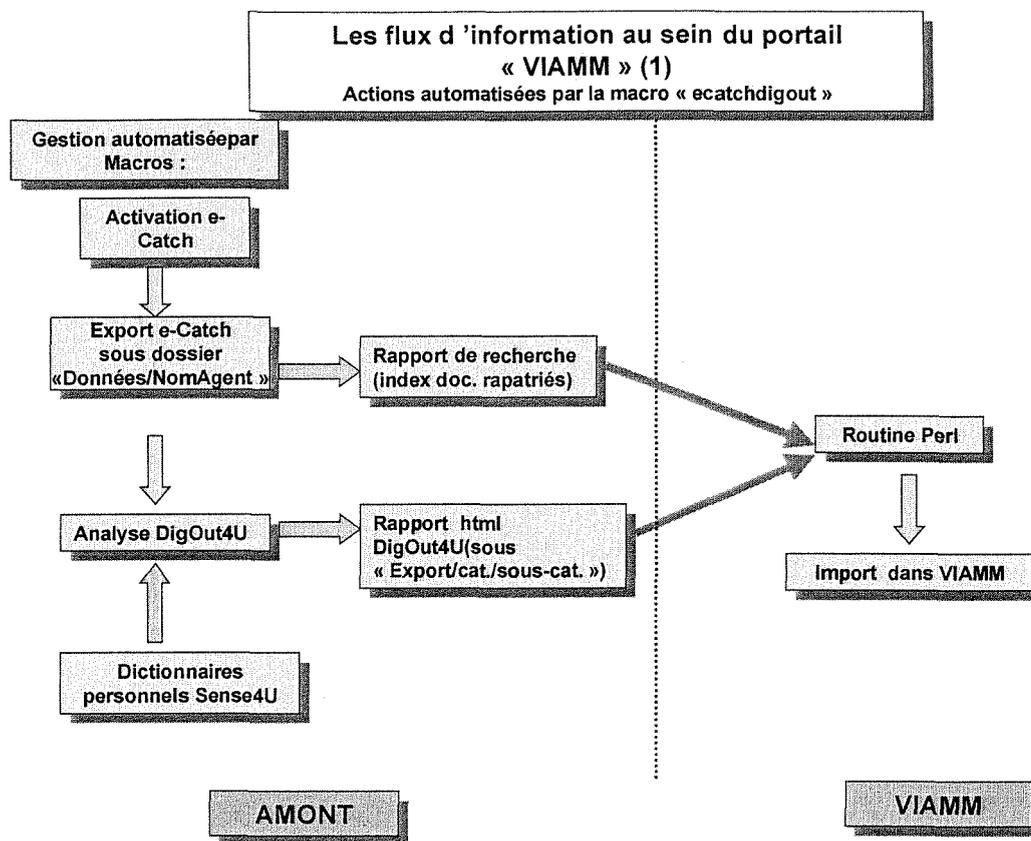


Figure 23 : Flux d'information provenant de l'outil e-Catch

Un autre exemple de macro concerne l'outil *Copernic* et ses agents spécialisés dans la presse. Une fois la recherche effectuée par Copernic (éventuellement en mode automatisé), la génération d'un rapport de synthèse est ordonnée par l'intermédiaire de la macro « copernic ». Le rapport comprend un certain nombre de données concernant les documents rapatriés. Une routine Perl permet là encore la conversion du rapport fourni par l'outil Copernic en un fichier avec l'extension .csv utilisable dans la base de données « VIAMM » (voir **Figure 24**).

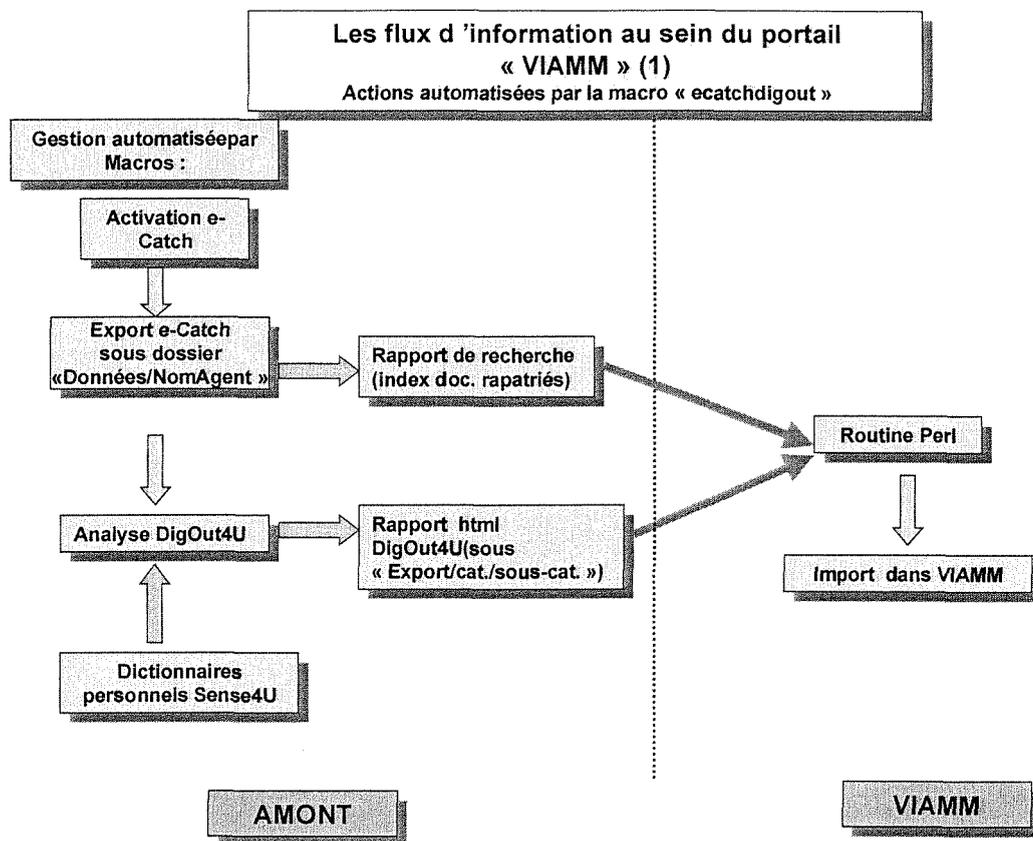


Figure 23 : Flux d'information provenant de l'outil e-Catch

Un autre exemple de macro concerne l'outil *Copernic* et ses agents spécialisés dans la presse. Une fois la recherche effectuée par Copernic (éventuellement en mode automatisé), la génération d'un rapport de synthèse est ordonnée par l'intermédiaire de la macro « copernic ». Le rapport comprend un certain nombre de données concernant les documents rapatriés. Une routine Perl permet là encore la conversion du rapport fourni par l'outil Copernic en un fichier avec l'extension .csv utilisable dans la base de données « VIAMM » (voir Figure 24).

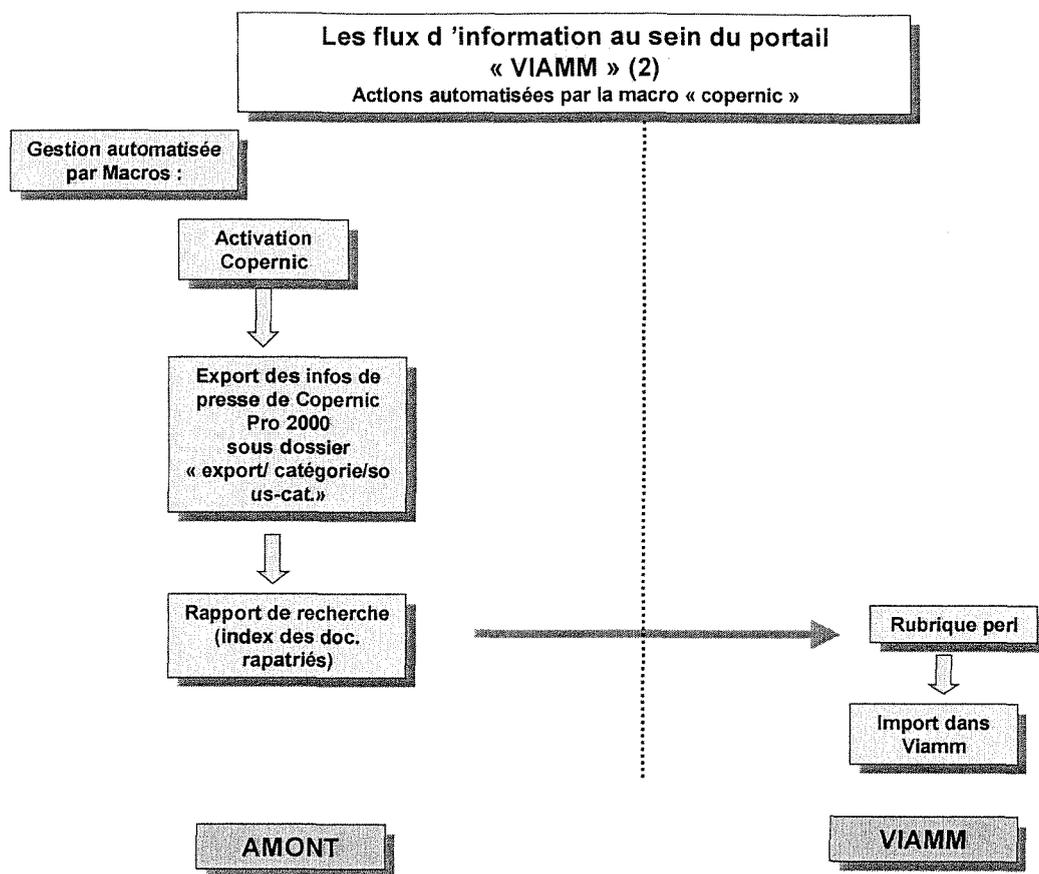


Figure 24 : Flux d'informations provenant de l'outil Copernic Pro 2000

En ce qui concerne la recherche de nouvelles adresses, une étape semi-automatisée est nécessaire :

- la sélection manuelle des adresses pertinentes parmi les résultats de recherche de Copernic
- le transfert des rapports de recherches fournis par « Copernic » vers le dossier adéquat situé sous un répertoire déclaré localhost pour le serveur Apache. Cette manipulation permet aux données rapatriées de faire l'objet d'un retraitement par e-Catch³, automatisé par la macro « ecatchexport ».

Cette manipulation semi-automatisée permet faciliter l'utilisation de l'outil e-Catch (voir : *Figure 25*).

³ En effet, le point de départ d'une nouvelle recherche e-Catch est obligatoirement une URL et pas un chemin local sur le poste de travail

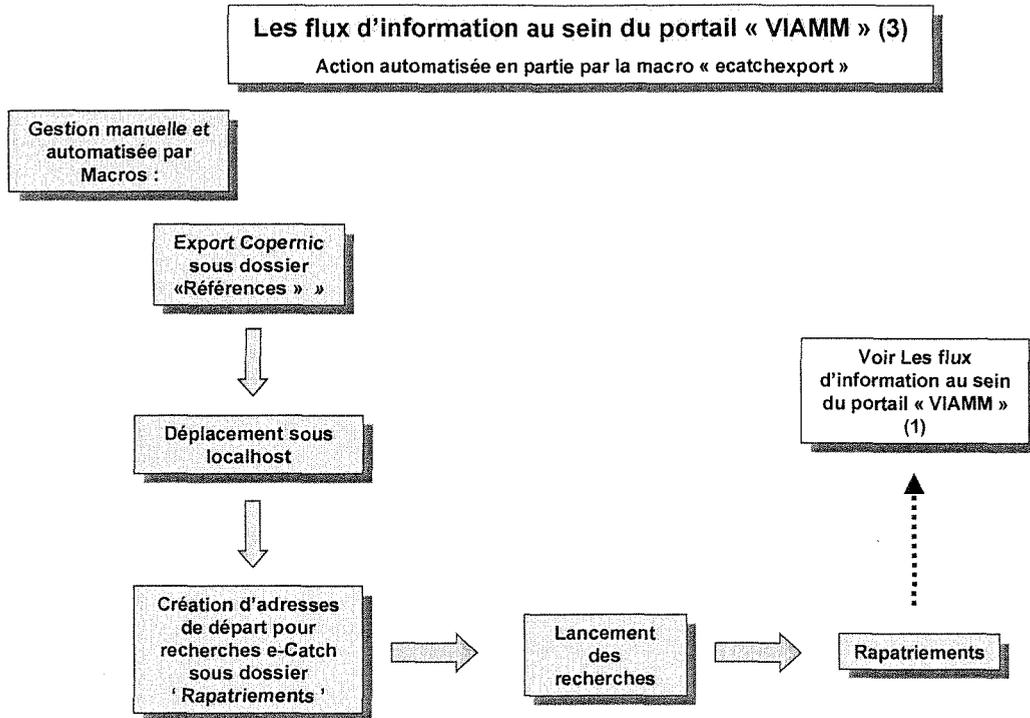


Figure 25 : Flux d'informations provenant de l'export « Copernic Thèmes » traité par e-Catch

IV.4. Une documentation interactive

Elle représente une part importante du projet et ceci, sur plusieurs plans. Tout d'abord, dans le cadre de la démarche Qualité, il est devenu indispensable de fournir des documents structurés fournissant les méthodes validées pour les outils employés en amont de Viamm. Cette documentation est un besoin côté client dans l'utilisation quotidienne du ou des outils fournis. Pour cela, il a été indispensable de fournir une documentation personnalisée au client. Tous les outils entrant dans le fonctionnement

général du portail ont été documentés en collant le plus possible aux protocoles d'utilisation définis au préalable en concertation avec les utilisateurs finaux.

Nous avons proposé d'intégrer la documentation dans le portail de deux façons :

1. une rubrique outils pointe sur les différents paragraphes des documentations des outils à partir d'une grille comparative. Cette grille est jointe en annexe (). Voir également la .

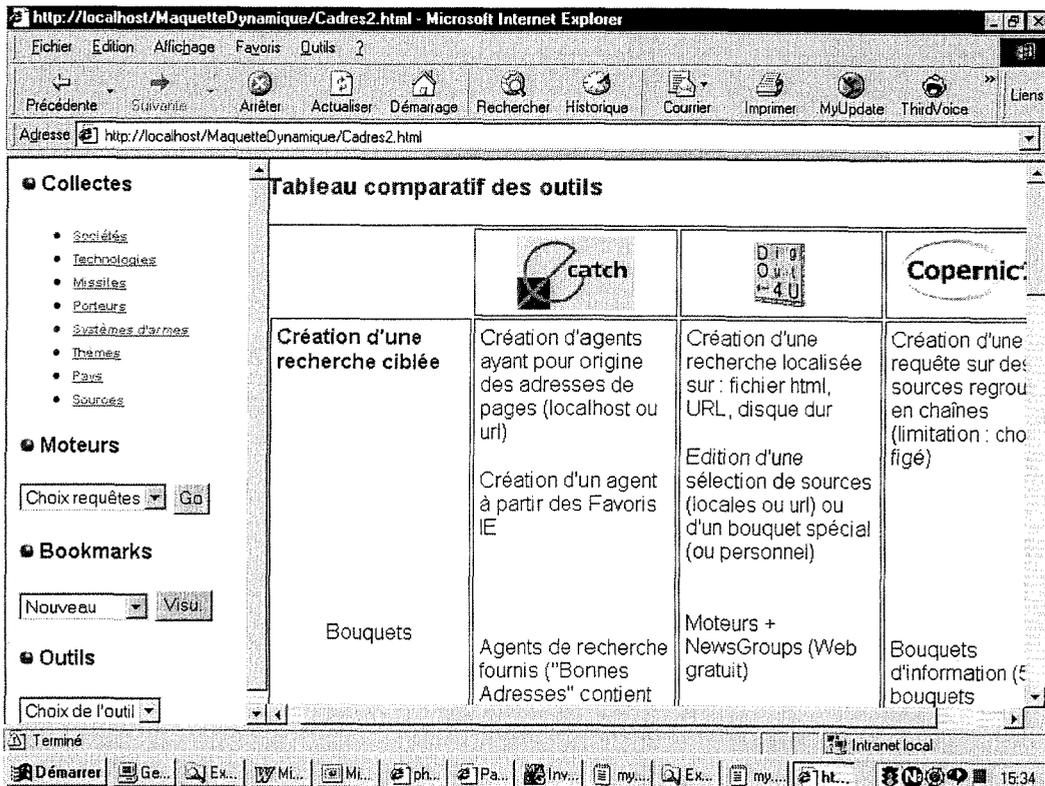


Figure 26 : Tableau comparatif de la documentation en ligne des outils amonts

2. Des messages déroulants et des "post-it" sont insérés dans les pages du portail. Ils rappellent les grands principes ou les particularités de fonctionnement des pages. Les textes défilants fonctionnent en continu et informent l'utilisateur sur les actions à effectuer. Les post-it sont activables et peuvent être refermés : ils vont plus loin dans les précisions et explications sur le fonctionnement du portail Viamm.

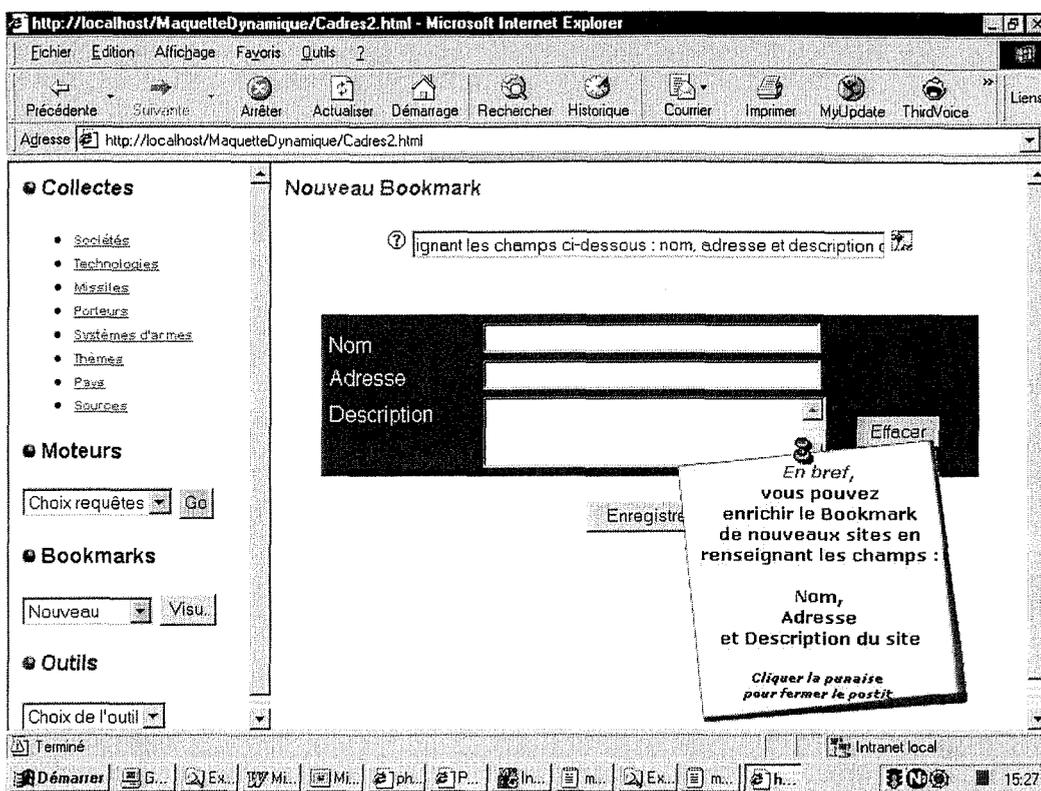


Figure 27: Exemple de page comprenant une aide en ligne sous forme de texte défilant et de Post-It

V. VIAMM en pratique

V.1. Scénarios d'utilisation

Les scénarios d'utilisation sont au nombre de quatre :

Scénario 1 : "Adresses validées à surveiller"

1. Paramétrage de l'outil e-Catch sur des adresses connues
2. Export des résultats
3. Reformatage des données à l'ouverture de VIAMM
4. Import dans Missile News

Scénario 2 : "Adresses qui présentent régulièrement de nouvelles informations à valider"

1. Paramétrage de l'outil e-Catch sur les adresses
2. Réanalyse des résultats rapatriés avec DigOut4U
3. Export des résultats et résumés
4. Sélection et reformatage des données dans VIAMM
5. Import dans Missile News

Scénario 3 : "Intégration massive de nouvelles informations"

1. Effectuer une requête lâche avec Copernic à partir de mots-clés simples (en mode manuel puis automatique suivant un calendrier à fixer)
2. Affiner cette recherche (option d'affinage automatisable)
3. Elimination des résultats a priori non pertinents puis export des résultats (étape manuelle ou semi-automatique en utilisant DigOut4U)
4. Surveillance et aspiration régulière des adresses avec e-Catch à partir des résultats exportés par Copernic. Intégration éventuelle de nouvelles adresses en utilisant la fonction "Lister les adresses identiques d'e-Catch"
5. Retour à l'étape 2 des scénarios 1 et 2 suivant le cas.

Scénario 3 : "Intégration progressive de nouvelles informations"

1. Effectuer une requête précise avec un moteur de recherche dédié à partir de mots-clés simples
2. Lecture rapide des résultats les plus pertinents (score de pertinence élevé)
3. Indiquer dans la rubrique Bookmarks de VIAMM les nouvelles adresses avec la catégorie « En Test »
4. Surveillance et aspiration régulière des adresses avec e-catch

V.1.1. Mises au points à réaliser avec les différents outils

Pour les scénarios ci-dessus, la mise au point nécessaire avec les outils amont se limite aux manipulations suivantes :

e-Catch	<ul style="list-style-type: none"> • Aspiration profonde depuis 1 adresse • recherche des pages pertinentes: → fonction de recherche • paramétrer les calendriers → tester la comparaison des pages pour ajuster la fréquence de rafraîchissement • tester la fonction d 'export
Copernic	<ul style="list-style-type: none"> • Recherche sur mots clés / chaînes • affinage sur une recherche au moment de la duplication (⇒ téléchargement dans 1 répertoire par défaut) • consulter les vues • affinage sur les sélections • automatisation du raffinage sur l 'agent initial • programmation de l 'onglet veille
DigOUt4U	<ul style="list-style-type: none"> • Recherche sur un sous- répertoire de <i>Rapatriements</i> • amélioration du dico personnel avec Sense4U • vérifier les résumés obtenus (dans <i>Export</i>) → ajouter des mots associés dans le dico personnel

V.2. Collecte des retours utilisateurs

La gestion des retours utilisateurs est un aspect essentiel du projet. S'agissant d'une maquette de portail, celle-ci est prévue pour durer le temps d'aider le client à définir très précisément ses besoins au fur et à mesure des usages. De nouvelles versions sont prévues fin 2001, à moins de trouver sur le marché un produit adéquat, ce qui n'est pas le cas aujourd'hui.

Afin de recueillir les avis des utilisateurs à distance et sous une forme plus objective que les entretiens réalisés a posteriori, le portail Viamm intègre un système de collecte des avis des utilisateurs sous la forme de pages interactives où chaque utilisateur peut donner son avis et exprimer s'il est d'accord avec d'autres avis. Ce système est une adaptation du script Perl "Cliff's Voting Booth Version 1.5".

Collecte des retours utilisateur portail Viamm

Bienvenu ! Vous êtes invités à donner votre avis sur ces sujets, ou [ajouter votre propre sujet](#)

rubrique Bookmarks	Vote!	resultats
rubrique Moteurs	Vote!	resultats
catégories	Vote!	resultats
résumés	Vote!	resultats
rubrique Outils	Vote!	resultats

Ajout d'un sujet:

Posez votre sujet (problème, thème, question, ...):

Saisir un choix: _____

Figure 28 : Extrait du formulaire de vote (choix du thème de vote)

A partir de cette page d'accueil, l'utilisateur choisit un sujet et vote, en ajoutant éventuellement son avis complémentaire comme illustré dans la figure

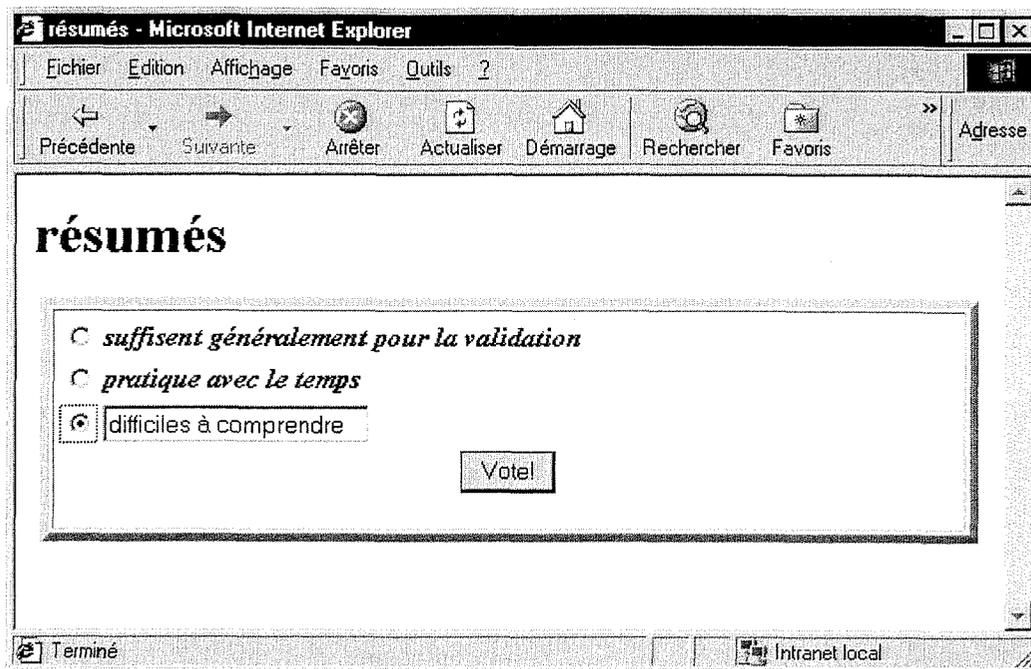


Figure 29 : Extrait du formulaire de vote (réponse en texte libre)

Des sujets clés en relation avec le fonctionnement du portail ont été prédéfinis afin de canaliser certaines remarques et pour inciter les utilisateur à en ajouter. Contrairement à un forum le regroupement des avis relatifs à un même sujet permet de prendre rapidement connaissance de ce qui est déjà connu. L'utilisateur doute moins de la pertinence de son avis.

Il n'était pas possible d'initialiser un service client dans le cadre de ce projet. Pour limiter les appels directs tout en permettant au client de visualiser l'"état" de leur demandes implicites ou la prise en compte de leur avis, un mode *Résultat* donne un aperçu des réactions des utilisateurs sur un sujet préalablement exprimé.

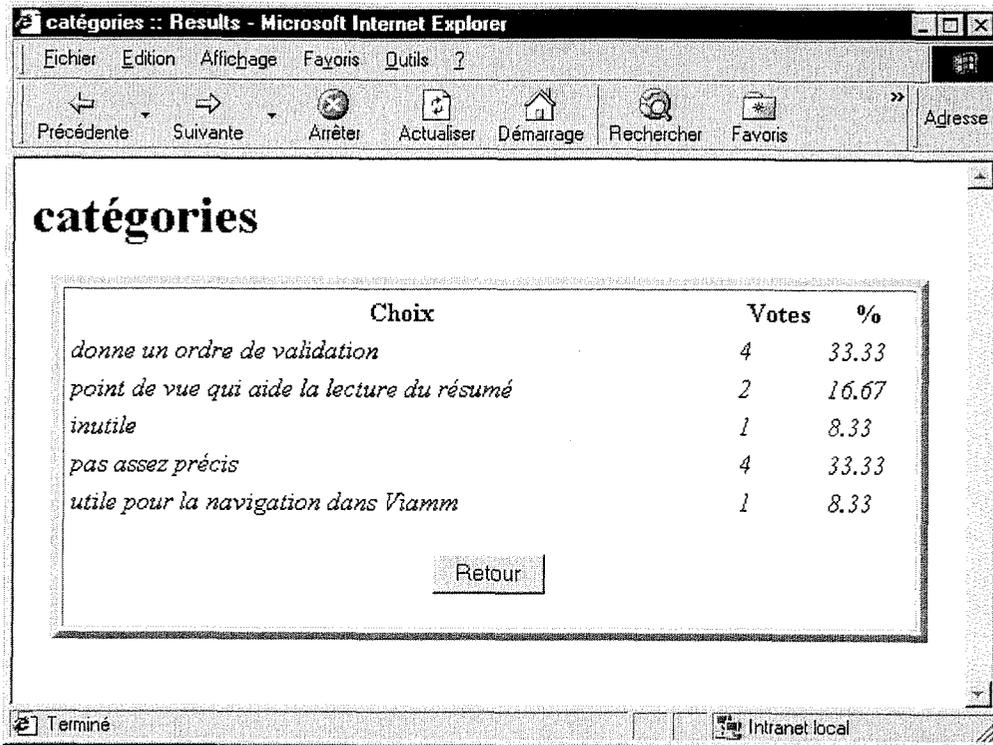
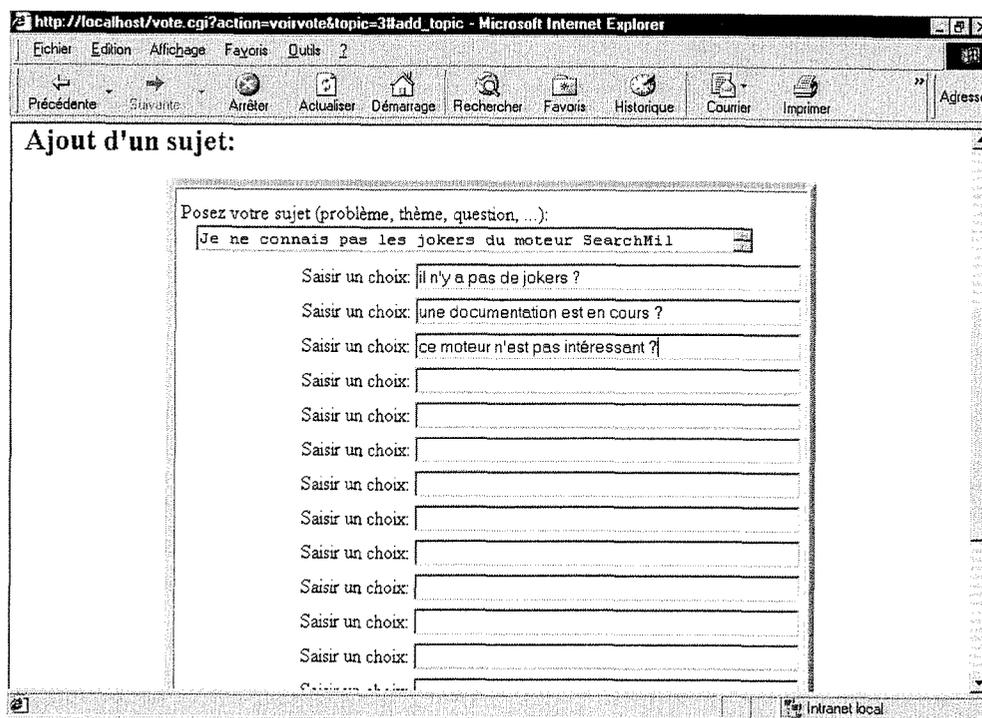


Figure 30 : Extrait du formulaire de vote (visualisation des résultats)

Pour les problèmes plus urgents, l'utilisateur peut utiliser le système de collecte des retours utilisateurs comme un forum où exprimer son problème en attendant qu'une personne lui réponde.



The screenshot shows a Microsoft Internet Explorer browser window with the address bar displaying `http://localhost/vote.cgi?action=voivote&topic=3#add_topic`. The browser's menu bar includes 'Fichier', 'Edition', 'Affichage', 'Favoris', and 'Outils'. The toolbar contains icons for 'Précédente', 'Suivante', 'Arrêter', 'Actualiser', 'Démarrage', 'Rechercher', 'Favoris', 'Historique', 'Courrier', 'Imprimer', and 'Adresse'. The main content area is titled 'Ajout d'un sujet:' and contains a form with the following elements:

- A text input field with the placeholder text 'Posez votre sujet (problème, thème, question, ...):' and the entered text 'Je ne connais pas les jokers du moteur SearchMil'.
- A series of radio button options, each preceded by the text 'Saisir un choix:':
 - il n'y a pas de jokers ?
 - une documentation est en cours ?
 - ce moteur n'est pas intéressant ?
 - [Empty]
 - [Empty]

Figure 31 : Extrait du formulaire de vote (création de nouveaux sujets)

Il s'agit là d'une début de système de résolution collaborative d'incident. Pour améliorer la qualité des réponses ou proposer d'autres alternatives, un mode administrateur est prévu pour gérer les avis déjà édité.

Ce mode permet d'effacer éventuellement certaines réponses hors propos, de regrouper certains avis, ou de supprimer un sujet qui correspond à une demande implicite et d'ajouter en nouveau sujet la solution.

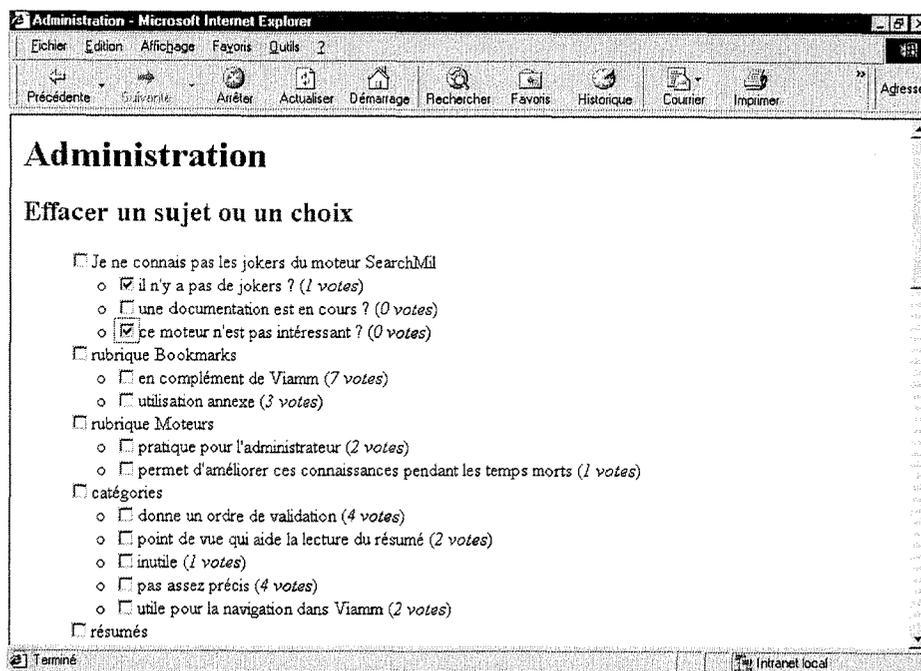


Figure 32 : Extrait du formulaire de vote (zone réservée à l'administrateur du portail Viamm)

VI. Bilan

VI.1. Bilan technique

L'élaboration de ce projet de stage m'a permis d'aborder, en plus de réutiliser un certain nombre de connaissances acquises au cours de l'année de formation (HTML, gestion de bases de données, langage SQL, Internet, Recherche Documentaire Spécialisée...), des langages nouveaux, de découvrir la gestion de bases de données compatible avec le Web : MySQL. En effet, la gestion de la base de données nécessitait une certaine connaissance du langage PHP que je ne maîtrisais pas du tout avant le début du stage. L'interrogation de la base de données MySQL par des requêtes SQL se fait à partir du langage PHP.

Il m'a aussi permis de m'initier à la manipulation de fichiers de configuration tels que ceux du serveur Apache qu'il a fallu modifier pour faire cohabiter l'ensemble des technologies sous ce serveur sur une même station de travail.

L'initiation au langage Perl m'a permis de découvrir la mise en place de routines qui, dans VIAMM, permettent notamment d'effectuer des traitements sur les fichiers pour leur permettre de devenir compatibles avec la base Missiles News.

L'utilisation d'un outil tel que WinMacro m'a permis de me familiariser avec la construction de macros qui permettent l'automatisation de certaines tâches répétitives qui seraient dévolues sinon aux utilisateurs du portail VIAMM.

La création de fichiers batch a également été nécessaire automatiser certains lancements d'outils.

Dans cette énumération de nouveautés, il ne faut pas oublier la prise en main des outils amont (e-Catch, DigOut4U et Copernic) dont il a fallu tirer le meilleur profit en comparant leurs capacités propres pour alimenter le portail Viamm en documents

nouveaux. Des travaux d'évaluation ont été nécessaires incluant une vision critique des produits existants et la rédaction de documentations complètes sur chacun des outils.

Ce travail s'est matérialisé par la rédaction de documentations spécifiques à chacun des outils de veille (Digout4U, Copernic 2000 Pro et e-Catch) pour aider les utilisateurs du portail Viamm à paramétrer leurs recherches.

VI.2. Bilan Projet

Une expérience nouvelle pour moi a été d'intégrer une équipe avec un projet commun et de participer à toutes les phases de l'élaboration du produit jusqu'à l'installation finale chez le client. La participation à des rencontres professionnelles m'a permis de mieux appréhender le sujet.

Sur le plan de la gestion du projet, j'ai pu mettre en pratique le savoir-faire acquis tout au long de l'année comme la formation sur la conception de serveurs Web, la formation en qualité, la recherche documentaire spécialisée...

Le contact avec les utilisateurs finaux du portail a été constant. Tout au long du projet, des réunions concernant son évolution ont été organisées nécessitant la rédaction de comptes-rendus. Au cours de ces réunions, j'ai appris à effectuer des démonstrations de l'outil, à interpréter un retour client. La relation directe avec le client est un plus indéniable.

De plus, dans le cadre de la démarche de la société, j'ai eu l'occasion de participer à la rédaction de documents Qualité (ISO 9001).

Enfin, la gestion du projet dans le temps a été également très importante et on peut souligner l'aide que peut apporter un logiciel comme MS Project.

VII. Bibliographie

[BAS1999]**Bases Publications**, 14 Outils pour une veille efficace sur le Net, *NetSources, hors-série no 2*, 1999.

[CHA1999]**Chaleat P., Charnay D.**, Programmation HTML et JavaScript, *Eyrolles, quatrième tirage*, 1999.

[DEZ2000]**Dézéus J.P.**, PhpInfo.Net - Site d'information PHP / MySQL, <http://www.phpinfo.net>, 2000.

[GAR2000]**Gardarin G.**, Internet/Intranet et bases de données, *Eyrolles, deuxième tirage*, 1999.

[GUT1999]**Guthrie P.**, Comprendre et exploiter une base de données, *NetPro, no 24*, Juillet-Août 2000, pp. 59-64.

[HER1999]**Herellier J.M., Merigod P.**, Pages Web dynamiques avec ASP-PHP-SQL, *Campus Press*, 1999.

[PHP2000]**PHP France**, Site d'informations PHP - PHP France, <http://phpfrance.com/news>, 2000.

[ROU&BAY 2000a] **Roussel D., Bayon N.**, Documentation e-Catch 3.0, *EADS*, 2000.

[ROU&BAY 2000b] **Roussel D., Bayon N.**, Documentation Copernic 2000 Pro, *EADS*, 2000.

[ROU&BAY 2000c] **Roussel D., Bayon N.**, Documentation DigOut4U, *EADS*, 2000.

Annexes

Table des annexes

<i>Annexes</i>	1
<i>Table des annexes</i>	2
I.1. Documentation du code	3
I.2. Fonctionnement PHP	5
I.3. Grille comparative des outils de veille	7

I.1. Documentation du code

Un outil de génération automatique de documentation a permis tout au long du développement du portail Viamm, tout en documentant le code source, de construire une documentation HTML issue du traitement de ce code source par l'outil SDD. S'agissant d'un démonstrateur dont le code peut être maîtrisé par une personne, la description du code vaut pour la description des unités logicielles. Une description intermédiaire est générée à partir des commentaires insérés dans le code par l'utilitaire `esd2html.pl`. Le résultat se trouve dans un répertoire `doc` situé sous le répertoire principal du démonstrateur.

Cet outil est particulièrement intéressant car l'administrateur du portail Viamm peut avoir des modifications à effectuer directement dans le code source. Ainsi, grâce à cette documentation automatique, les manipulations à effectuer ont pu être documentées à même le code source (introduction de commandes qui n'ont plus qu'à être décommentées pour fonctionner).

Une exemple de documentation est montré Figure 1.

I.2. Fonctionnement PHP

Actuellement, il existe deux versions du langage : PHP 3 (version 3.016) qui est un langage interprété pur, ainsi que PHP 4 (version 4.0. RC1), sous forme de compilateur de code objet. Ces deux versions peuvent être utilisées soit comme n'importe quel CGI portable soit comme module intégré pour un large éventail de serveurs Web parmi lesquels se trouve celui utilisé pour le projet : le serveur Apache.

Le code PHP fait partie intégrante de la page HTML. Le code sous-jacent reste totalement inconnu pour l'utilisateur arrivant sur une telle page. Ce code est interprété côté serveur avant d'être traduit en HTML.

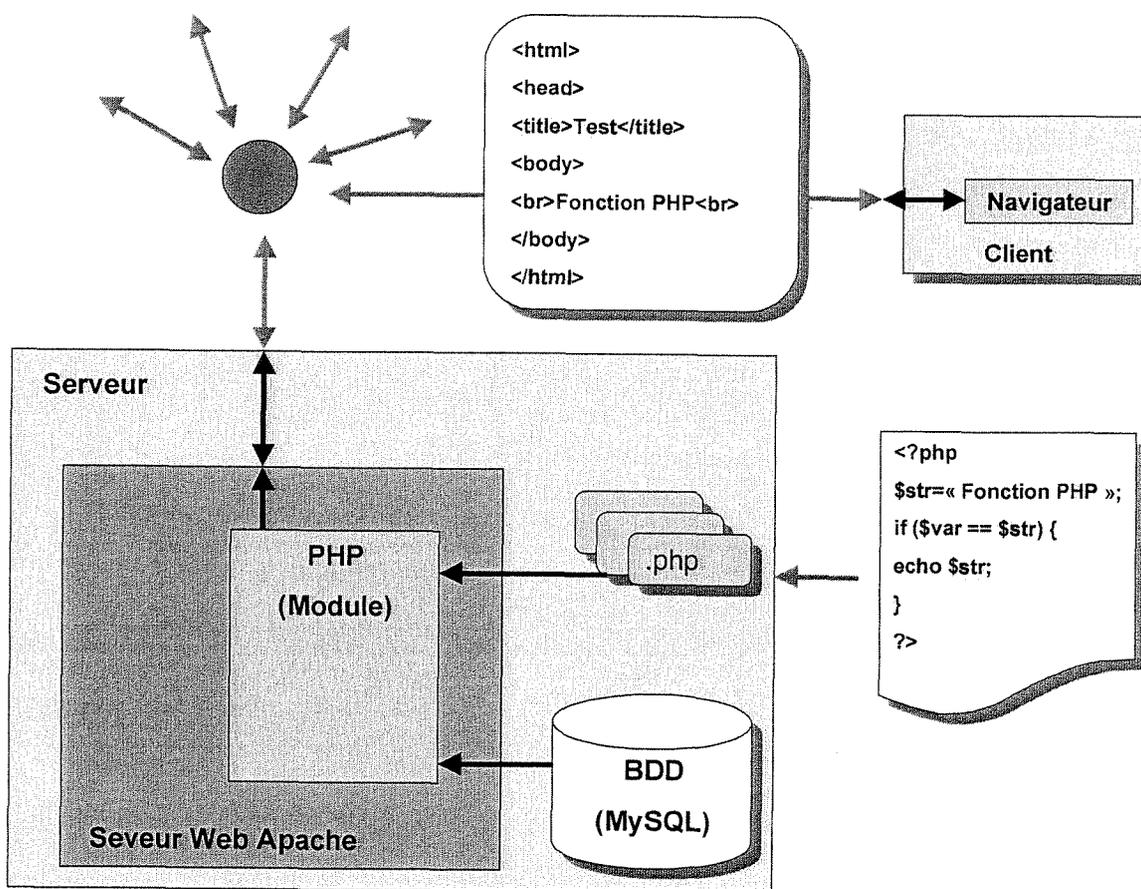


Figure 2 : Schéma présentant le fonctionnement général du langage PHP intégré au langage HTML

Pour ce faire, le serveur Web lance l'interpréteur PHP qui traduit le document demandé et exécute le code source de la page (voir Figure ci dessous). Les commandes figurant dans la page sont interprétées et le résultat prend la forme d'une page HTML publiée à la place du code source dans le même document. A l'issue de cette phase de traduction, la page modifiée est envoyée au client pour y être affichée par le navigateur.

I.3. Grille comparative des outils de veille

	e-Catch 3.0	DigOut4U	Copernic Pro 2000
Création d'une recherches ciblées Bouquets Langues	Création d'agents ayant pour origine des adresses de pages (localhost ou url) Création d'un agent à partir des Favoris IE Agents de recherche fournis ("Bonnes Adresses" contient une rubrique magazine et quotidiens) -	Création d'une recherche localisée sur : fichier html, URL, disque dur Edition d'une sélection de sources (locales ou url) ou d'un bouquet spécial (ou personnel) Moteurs + NewsGroups (Web gratuit) Multilinguisme (traduction des mots clés ou limitation à une langue donnée)	Création d'une requête sur des sources regroupées en chaînes (limitation : choix figé) Bouquets d'information (50 bouquets prédéterminées pas d'ajout possible) Limitation à une langue donnée
Critères de recherche spécifiques	Recherche à partir d'une adresse Gestion de la profondeur d'exploration Gestion des mots de passe Gestion de certains champs de formulaires	Construction de requête Limitation de la profondeur d'exploration Serveurs interdits Calcul de pertinence à partir d'un thésaurus multilingue	Trois types de recherches Recherches élaborées (affinage des résultats par des opérateurs booléens et troncature implicites. Recalcul du score de pertinence)
Rapatriement Calendrier Dossiers	Calendrier de rafraîchissement des pages téléchargées + (voir création d'agents : chacun peut regrouper plusieurs pages)	- -	Options de téléchargements Calendrier (pour différer la recherche) Dossiers (regroupant plusieurs recherches)
Résultats	Fonctions de recherche avancée (dans code ou texte)	Statistiques sur les serveurs parcourus Options de filtrage	Sélection d'un sous-ensemble de documents Tri des résultats
Alerte e-mail	-	-	Alerte
Signalement visuel au	Identification des	Visualisation de la	

démarrage	modifications de sites (couleur de l'icône de Netscape)	pertinence des documents (couleur de l'icône)	
Comparaison des modifications	Mode Comparaison (icônes eCatch entourant modif.)	-	-
Rapports	-	Sur les recherches (voir statistiques sur les serveurs)	Rapports sur les recherches (fonction de consultation) Rapport sur les tâches de veille
Export des données Qualité de l'export	Fonction d'export Fichier "index" Options d'export : format html ou texte	Fonction d'export Fichier "Résultat de recherche"	Création d'un bookwork à partir d'une liste de documents sélectionnés (avec possibilités de mise-à-jour ou de filtrage des nouveautés)
Performances Elimination des doublons Vérification des liens			

