HAL
open science

Mise en place de domaines scientifiques dans l'observatoire ARTIST

Djamila Safa

To cite this version:

Djamila Safa. Mise en place de domaines scientifiques dans l'observatoire ARTIST. domain_shs.info.comm. 2005. mem_00000259

HAL Id: mem_00000259

https://memsic.ccsd.cnrs.fr/mem_00000259

Submitted on 14 Oct 2005

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Henri Poincaré
Université Nancy 2
Institut National Polytechnique de Lorraine
DESS Information Scientifique et Technique - Intelligence Economique

Promotion 2004-2005

RAPPORT DE STAGE

Mise en place de domaines scientifiques dans l'observatoire ARTIST

Stage effectué du 9 mai au 9 septembre 2005
A l'Institut de l'Information Scientifique et Technique **(INIST-CNRS)**

Par
Djamila SAFA

Maître de Stage
Jacques DUCLOY
INIST-CNRS
54514 Vandœuvre lès Nancy
03.83.50.20.00

Stage soutenu le : 23 septembre 2005

SOMMAIRE

AVANT PROPOS **4**
REMERCIEMENTS **5**
INTRODUCTION **6**

1 PRESENTATION DE L'INIST-CNRS ET DU PROJET ARTIST **7**
1.1 INIST : ORGANISME EN PLEINE MUTATION **7**
1.1.1 Rôle et statut *7*
1.1.2 Création d'une DIS au CNRS *7*
1.2 MISSIONS ET SERVICES **8**
1.2.1 Fonds documentaires *8*
1.2.2 Bases de données et portails *8*
1.2.3 Service de veille scientifique et technologique *9*
1.3 DEPARTEMENT PROSPECTIVES ET RELATIONS EXTERIEURES **9**
1.3.1 Missions et organisation *9*
1.3.2 Activités éditoriales autour du web *10*
1.4 PROJET ARTIST **11**
1.4.1 Initiative ARTIST *11*
1.4.2 Mise en place *12*
1.4.3 Volenté de partenariat *12*

2 PRESENTATION DES MISSIONS DU STAGE **13**
2.1 CONTEXTE ET OBJETCTIFS **13**
2.2 SITE ARTIST ET REVUES SCIENTIFIQUES **14**
2.3 SITE ARTIST ET FORUMS **15**
2.2 SITE ARTIST ET OBSERVATOIRE **16**
2.2.1 objectifs de l'observatoire *16*
2.2.2 organisation de l'observatoire *16*

3 ADAPTATION EDITORIALE ET CONTRIBUTION A LA STRUCTURARTION DU SITE ARTIST **17**
3.1 SPIP OUTIL D'EDITION COLLABORATIVE EN LIGNE **17**
3.1.1 Définition *17*
3.1.2 Historique *17*
3.1.3 Architecture d'un site SPIP *17*
3.1.4 Caractéristiques de SPIP *19*
3.2 CONCEPTION DE FICHES D'INFORMATION **19**
3.2.1 Définition et réalisation *19*
3.2.2 Difficultés rencontrées *20*
3.3 ARCHIVAGE DANS UNE VISION ENCYCLOPEDIQUE **21**
3.3.1 Indexation *21*
3.3.2 Difficultés rencontrées *22*

4 VEILLE SUR LE DOMAINE BIOMEDICAL ET L'APPROPRIATION DES TECHNOLOGIES DE L'IST 25

4.1 DEFINITION DE LA STRATEGIE DE RECHERCHE 25

4.2 RECHERCHE ET COLLECTE 25

4.2.1 construire une requête 26

4.3 TRAITEMENT ET ANALYSE 27

4.3.1 outils et données 27

4.3.2 procédures d'analyse 28

4.4 RESULTATS D'ANALYSE 33

4.4.1 Résultats obtenus avec Miriad à partir de la base PASCAL 33

4.4.2 Résultats obtenus avec Leximine à partir des bases PASCAL et Pubmed 35

4.4.3 Résultats d'analyse obtenus sur Keywatch 45

4.4.4 Difficultés recontrées 48

4.5 DIFFUSION DES RESULTATS 48

CONCLUSION ET PERSPECTIVES 51

ABREVIATIONS 53

GLOSSAIRE 54

REFERENCES BIBLIOGRAPHIQUES 57

ANNEXES 59

LISTE DES FIGURES ET TABLEAUX 73

Avant propos

Ce stage a été réalisé dans le cadre du DESS « Information Scientifique et Technique-Intelligence Economique » cohabilité par les trois universités de Nancy (Université Henri Poincaré Nancy 1, Université Nancy 2, Institut National Polytechnique de Lorraine).

Dans le cadre de ma formation de DESS IST-IE, j'ai eu la possibilité d'effectuer mon stage de fin d'année d'études universitaires à l'INIST - CNRS (INstitut de l'Information Scientifique et Technique - Centre National de la Recherche Scientifique). Cet institut de renommée internationale dans le domaine de l'Information Scientifique et Technique (IST), collecte, traite l'information issue de la recherche scientifique et la diffuse sous forme de copies de documents scientifiques ou de notices bibliographiques. Il propose également des produits et services documentaires sous forme électronique via Internet. Ses différents portails participent au développement de la diffusion de l'information. L'objectif de ce stage était de mettre en place des domaines scientifiques dans l'observatoire ARTIST.

Remerciements

Je tiens à remercier :

- Monsieur Jacques DUCLOY pour la confiance qu'il m'a accordée, son suivi pendant la réalisation de mon stage
- Monsieur Philippe HOUDRY pour son aide et tous ses conseils
- Madame Marie-Catherine GUNET pour avoir relu mon rapport
- Hélène CREUSOT pour sa collaboration et son soutien
- Jean-François NOMINE pour son accueil et sympathie
- Solveig VIDAL pour son aide précieuse
- Tous les membres du Département Produits et Services, Département Prospectives et Relations Extérieures et du Service veille pour leur accueil et leur sympathie.

J'aimerais dédier ce travail à :

- La mémoire de : mon très cher père Arezki, ma très chère sœur Nadia et mes chers grands-parents pour leur éternelle présence
- Ma très chère mère Fatima pour ses encouragements et sa confiance en moi
- Mon grand frère Hassène, tout d'abord parce qu'il a fait naître l'envie de me lancer dans cette aventure, puis parce qu'il a ravivé et stimulé cette envie en me rassurant dans les moments de doute, de peur ou d'incertitude. Son attention, ses conseils, ses paroles, sa protection sont des trésors inestimables.
- Mes frères et sœurs
- Toute ma famille et mes ami(e)s

Introduction

L'Information Scientifique et Technique, souvent abrégée en "IST", est diffusée sous forme d'articles, actes de congrès, thèses, ouvrages, brevets et rapports. Elle est généralement structurée et identifiable au travers des catalogues de bibliothèques, des bases de données bibliographiques et en texte intégral, qui recensent et diffusent la littérature scientifique mondiale. [1][12]

Dans le contexte d'une concentration du marché et de l'apparition de monopoles, l'IST relève essentiellement de sociétés commerciales. Dans ce cadre l'INIST a mis en place un site Web, **ARTIST** (**A**ppropriation par la **R**echerche des **T**echnologies de l'**IST)** : http://artist.inist.fr) constitué d'un site public à caractère éditorial, géré par SPIP et un observatoire qui doit recenser et promouvoir les outils, normes, organismes et projets ayant une valeur exemplaire par rapport à cette logique d'appropriation.

Mon stage s'est déroulé à l'INIST, au DPRE (Département Prospective et Relations Extérieures), en relation avec le Service veille. L'objectif de mon travail a été la mise en place de domaines scientifiques dans l'observatoire ARTIST. Dans un premier temps, je présenterai l'organisme d'accueil : l'INIST. Dans un second temps, j'expliquerai toutes les étapes du travail réalisé : la publication en ligne et l'étude de veille sur le domaine biomédical et son appropriation des technologies de l'IST. Je terminerai par un bilan exhaustif des missions qui m'ont été confiées en y incluant une perspective d'avenir pour ce nouveau produit de l'INIST issu de l'initiative de M. Jacques Ducloy - conseiller scientifique du directeur de l'INIST- et de ses collaborateurs.

1. Présentation de l'INIST-CNRS et du projet ARTIST

1.1. INIST : organisme en pleine mutation

1.1.1. Rôle et statut

L'Institut de l'Information Scientifique et Technique (INIST) est une unité propre de service (UPS) du Centre National de la Recherche Scientifique (CNRS) chargée de collecter, analyser et diffuser les résultats de la recherche internationale. Voici quelques informations sur cet organisme extraites du site (http://www.inist.fr/actu/inist_cest.php) :

L'INIST met à disposition des grands organismes publics de recherche, de l'enseignement supérieur et du milieu socioéconomique, un ensemble de ressources et de services pour accéder à l'information scientifique et technique internationale.

Fournisseur de copies de documents, producteur de bases de données multilingues et multidisciplinaires recensant l'essentiel de la littérature internationale avec plus de 17 millions de références dans la plupart des domaines de la recherche, l'INIST propose également une gamme complète de services d'accès à l'information scientifique et technique accessibles sur Internet et adaptés aux besoins de l'utilisateur. [9]

1.1.2. Création d'une DIS au CNRS

La politique du CNRS en matière d'IST est en pleine évolution. Une Direction de l'Information Scientifique (DIS) vient d'être créée ; la décision en a été officiellement publiée dans **SG hebdo** (La publication du secrétariat général du CNRS destinée aux unités de recherche et de service) - n°71 de 8 juillet 2005 :

« Par décision du directeur général du CNRS, Bernard Larrouturou, une direction de l'information scientifique (DIS), rattachée à la direction scientifique générale, est créée à compter du 1er juillet 2005 au CNRS. Laurent Romary, directeur de recherche à l'INRIA, est nommé directeur de la DIS à compter de ce même jour.

La direction de l'information scientifique a pour missions d'élaborer et de mettre en œuvre la politique du CNRS en matière d'information scientifique et technique, en lien avec les départements scientifiques et les instituts nationaux du CNRS, et en relation avec les autres acteurs du dispositif national de recherche ». [14]

La création de la DIS va avoir d'importantes répercussions sur l'organisme INIST. Pendant mon stage, j'ai eu de multiples discussions avec mes collègues au sein de

l'organisme, qui m'ont appris qu'il y aurait probablement des changements d'organisation au niveau du DPRE.

1.2. Missions et services

1.2.1. Fonds documentaires

L'INIST dispose d'un fonds documentaire unique en France couvrant l'essentiel de la littérature scientifique et technique internationale : sciences, technologie, médecine, sciences humaines et sociales. Plusieurs dizaines de millions de documents sont ainsi stockés sur les 27 km de rayonnage de ses magasins ou dans son Système d'Archivage Numérique (SAN). Ce fonds permet, à la demande, la reproduction papier ou numérique des documents ; il sert aussi à la constitution des bases de données produites par l'INIST. (Cf. figure 1) [10]

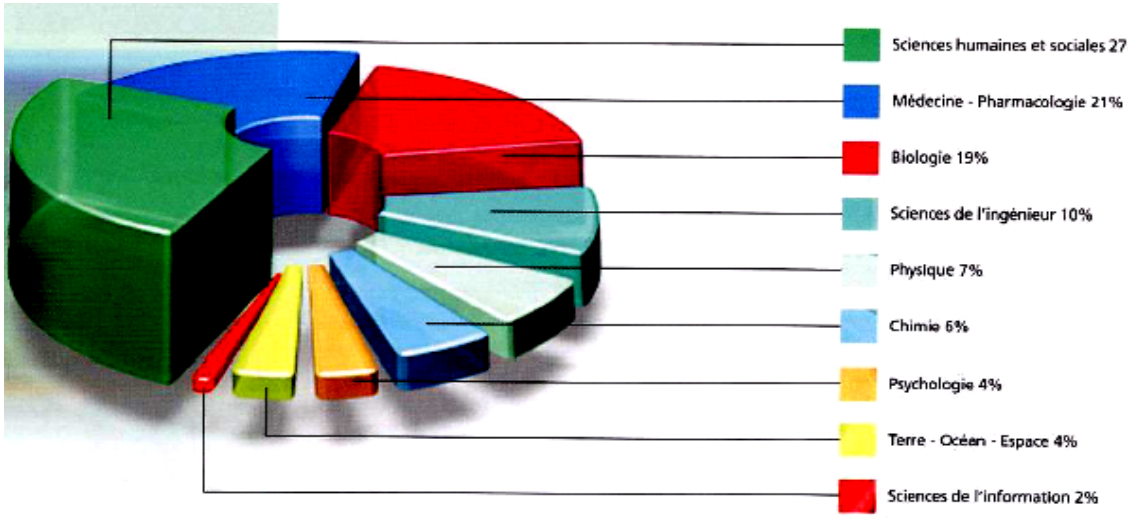


Figure 1 : *Répartition par domaines du fonds documentaire des périodiques (26000 titres) de l'INIST*

1.2.2. Bases de données et portails

L'INIST coopère avec des organismes en France et à l'étranger pour produire ses bases de données bibliographiques multilingues et multidisciplinaires : ***PASCAL**** Sciences, Technologie et Médecine, ***FRANCIS**** en Sciences Humaines et Sociales et ***ISD**** (International Science Database), couvrant l'essentiel de la littérature mondiale en science, technologie, médecine ainsi qu'en sciences humaines et sociales. [3] [8]

Les technologies de pointe utilisées en matière de stockage de l'information permettent à l'INIST d'offrir des services sur Internet (voir annexe 1), tels que :

- **ArticleSciences** : moteur de recherche et de commande en ligne d'articles scientifiques et techniques du fonds de l'INIST, associé à un module de paiement sécurisé en ligne.
- **BiblioSciences** : service actuellement réservé aux laboratoires du CNRS, BiblioSciences permet un accès mutualisé à l'information scientifique et technique regroupant plusieurs bases de données scientifiques.
- **ConnectSciences** : portail CNRS d'information scientifique et technique proposant, dans un environnement personnalisé, un accès unique et gratuit à toute une gamme de services spécialisés.
- **Services@inist** : l'ensemble des services documentaires de l'INIST comprenant :
  - ARTICLE@INIST, le catalogue des articles et monographies du fonds INIST depuis 1990.
  - FORM@INIST, un formulaire de commande en ligne de copies d'articles.
  - COMPTE@INIST, un module de suivi et de gestion des documents et des commandes de copies d'articles.
- **Thesa :** une base de données française gratuite qui signale les sujets de thèses en cours dans les Grandes Ecoles. Elle permet de connaître les thèmes de recherches menées par les laboratoires de Grandes Ecoles et d'identifier les différents acteurs de ces recherches. [10][11]

1.2.3. Service de veille scientifique et technologique

L'INIST conçoit, développe et utilise différentes technologies d'analyse de l'information grâce à des outils de veille technologique et documentaire tels que les plates-formes ***STANALYST****, ***DILIB****, **KeyWatch** ainsi que le logiciel **LEXIQUEST MINE** pour le traitement bibliométrique et l'analyse infométrique des données issues de diverses sources d'information, et en particulier des bases de données bibliographiques. [9][10]

Ces applications constituent une aide à l'élaboration de stratégies scientifiques, tant pour les chercheurs que pour les entreprises. Le Service veille m'a apporté son assistance pour réaliser mon étude de veille.

1.3. Département Prospective et Relations Extérieures de l'INIST / CNRS

1.3.1. Missions et organisation

Le DPRE, placé sous la responsabilité de Marc Guichard, est organisé en trois services (Service Prospective et Partenariats, Service Promotion-Communication et Service

Web-Traduction) (Voir annexe 2). Sa mission est d'assurer la communication institutionnelle de l'INIST ainsi que celle concernant les produits et services de l'INIST, de développer des partenariats au niveau national et international et de représenter l'INIST auprès des instances d'IST internationales.

Les services du DPRE permettent de fournir à l'INIST les éléments qui lui sont nécessaires pour établir sa stratégie à moyen et à long terme comme par exemple, de développer des partenariats au niveau national, européen et international en participant à la réalisation de projets tels que E-Biosci, IMIST, AEIA. Le DPRE assure aussi la coordinations de différents portails (ArticleSciences, ConnectSciences, PubliCNRS, TitaneSciences, BiblioVie, BiblioSHS, etc) (Voir annexe 1) pour lesquels ils négocient avec les éditeurs certains contrat de licence pour des revues électroniques et des bases de données. Il a également un rôle opérationnel dans le lancement de sites ***web**** et dans l'édition électronique, ce qui facilite l'accès en ligne à l'information, principalement pour les acteurs de la recherche publique (CNRS, Universités).

1.3.2. Activités éditoriales autour du web

La création, la diffusion et la mise à jour d'informations nécessaires à la politique de communication de l'Institut impliquent une présence sur le web et une dimension internationale. Un soutien logistique est donc assuré par les webmestres et une cellule technique au niveau du DPRE. Le site web de l'INIST et les espaces web associés permettent de promouvoir l'image de l'unité et de faire connaître son actualité, ses nouveaux produits et services. La communication Web s'adresse directement et dynamiquement aux clients mais aussi à tout public externe national et international intéressé par l'information scientifique et technique. Les principales activités web assurées par les services de DPRE sont :

- **le site web institutionnel :** les webmestres assurent la maintenance et la mise à jour du site Web institutionnel sous l'égide d'un comité éditorial.
- **les espaces web associés :** les webmestres créent des sites ou des espaces de communication Web liés à des opérations ponctuelles comme des manifestations, des partenariats, etc. (Exemple : BiblioInserm, Festival du film de chercheur).
- **le site sur le Libre Accès** (www.inist.fr/openaccess/) conçu pour informer la communauté scientifique sur les nouveaux modèles d'édition scientifique, ainsi que sur les différents mouvements et déclarations en faveur d'un accès libre et gratuit aux résultats de la recherche financée sur fonds publics

- **le site ARTIST** (http://artist.inist.fr/), mis en place récemment à l'INIST, a également bénéficié d'une coopération avec le DPRE.

1.4. Projet ARTIST

1.4.1. Initiative ARTIST

Le projet est né d'une réflexion sur le document numérique dont l'introduction bouleverse les pratiques de communication du monde de la recherche. L'explosion du document numérique modifie complètement le paysage de la communication scientifique et de l'IST. De façon paradoxale, au moment où la technologie se démocratise dans un contexte international, la vie de la recherche dépend de plus en plus des intérêts d'un petit nombre de groupes industriels de l'édition. Afin de favoriser un nouvel équilibre, une initiative éditoriale et collective nommée ARTIST (Appropriation par la Recherche des Technologies de l'IST) a été créée. [6]

L'objectif de cette initiative est de montrer comment l'appropriation des technologies de l'IST peut permettre à la recherche de reprendre la maîtrise de son système de diffusion des connaissances et de son mode de pilotage.

Pour favoriser l'appropriation des technologies de l'IST, ARTIST propose d'étudier et d'approfondir les thèmes suivants : multilinguisme, langues des sciences, infométrie, sciences de la vie et médecine, etc. Le projet ARTIST propose également d'organiser une activité de veille scientifique et technologique en utilisant le critère « appropriation » pour focaliser les études et organiser les résultats. Ce dispositif a pour but de repérer les expériences d'appropriation des technologies de l'IST en identifiant les problèmes et en analysant les résultats obtenus. Le dispositif est alimenté par des pratiques de veille et privilégie l'intérêt démonstratif et la possibilité de coopérer.

L'appropriation constitue le cœur de la problématique d'ARTIST. Les relations entre le monde de la recherche et celui de la communication scientifique (et plus généralement de l'IST) avaient trouvé un équilibre basé sur la séparation entre l'activité de recherche proprement dite et celles liées à la diffusion de la connaissance scientifique. Mais la donne a changé avec l'émergence de nouveaux modèles d'édition et de diffusion.

➢ **Autour du mot « appropriation »**

Le mot « appropriation » a un sens assez précis chez ceux qui l'emploient. Ils veulent dire d'abord que le public ou les citoyens doivent avoir un rôle actif à l'égard des technologies de l'information, qu'ils ne doivent pas être considérés uniquement comme une «cible» de leur diffusion, qu'enfin la démocratisation du numérique doit partir d'eux-mêmes, plutôt que l'inverse. [7]

L'appropriation des technologies de l'information et de la communication est un phénomène très complexe et elle représente un vrai défi. Mais, une chose est sûre : l'être humain doit toujours constituer le point de départ et le point d'arrivée de toute analyse, de toute mesure, de tout plan d'action. [2]

1.4.2. Mise en place

ARTIST est un dispositif à trois volets : un volet recherche, un volet veille scientifique et technologique et un volet éditorial. Il s'intéresse à un large éventail d'activités : édition électronique, constitution de bases de données, activités terminologiques, veille, ainsi qu'aux opérations innovantes dans lesquelles un mécanisme d'appropriation par la communauté scientifique est clairement identifié.

Ce projet a bénéficié d'une réflexion préalable et personnelle de quelques mois par les concepteurs de cette initiative. Celle-ci a été alimentée par de nombreuses discussions informelles dans des occasions assez diverses telles que la Semaine du Document Numérique, le colloque VSST, des contributions aux travaux de la mission IST du CNRS, etc.

Une toute petite maquette avait permis de concrétiser certaines idées. Elle est aujourd'hui, remplacée par un premier ensemble de services coopératifs en ligne (http://artist.inist.fr/). Deux listes de diffusion (artist@inist.fr : liste de diffusion pour un bulletin de type Flash et artist-gt@inist.fr : liste de discussion modérée pour les rédacteurs) [4]. D'autres listes constituent également un outil fédératif dans le site ARTIST (métadonnées, traduction, etc.).

1.4.3. Volonté de partenariats

ARTIST est en partenariat avec des organismes du monde de la recherche ou du transfert de technologie. Un premier noyau de comité scientifique et de pilotage a été mis en place (Voir annexe 3).

Concernant l'aspect observatoire, celui-ci s'inscrit naturellement dans le cadre des productions éditoriales et de veille technologique de l'INIST, au même titre que le « site sur le libre accès à l'information scientifique ». Cela dit, l'ensemble des acteurs de l'INIST se voit concerné en fonction de diverses actions telles que la mobilisation des ingénieurs sur la connaissance de l'innovation en IST dans leur domaine, le repérage d'acteurs en vue de coopération et l'implication possible de ceux-ci dans une activité éditoriale.

Concernant les aspects Recherche, la dynamique du document électronique est un élément clé de la problématique d'ARTIST et le ***RTP-DOC**** a démontré sa capacité à rassembler l'ensemble des acteurs concernés par la recherche sur ce thème. La présentation du projet ARTIST aux colloques ISD à Tunis, Isko France, VSST, au congrès CIDE à Beyrouth et aux journées de l'Ile Rousse ont permis de mieux cerner les principaux objectifs de ce projet (observatoire, revue…). [6] [4]

2.Présentation des missions du stage

2.1. Contexte et objectifs

ARTIST est une initiative sur l'appropriation par le monde de la recherche des technologies de l'IST et de la communication scientifique qui commence à trouver ses marques. C'est aussi un projet pilote pour de nouvelles pratiques éditoriales avec trois objectifs : forum pour chercheurs et praticiens, observatoire et revues scientifiques comme le montre la figure 2 (je vais décrire le site plus en détail dans la partie « Site ARTIST et forum » :

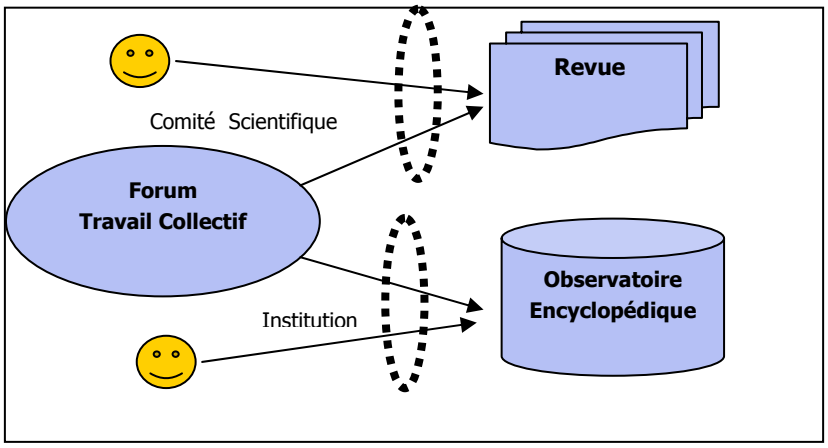


Figure 2 : *présentation général du site ARTIST*

Dans la pratique, ma mission consiste à mettre en place des services à plusieurs volets :

- Adaptations techniques et éditoriales pour établir des fiches d'information et de visualisation de résultats dans un esprit de communication scientifique. Les informations pérennes seront ensuite archivées dans une structure encyclopédique. Cette phase implique l'utilisation de l'outil SPIP, un système de publication en ligne.
- Veille sur le domaine biomédical et son appropriation des technologies de l'IST, en identifiant les références significatives d'appropriation dans ce domaine sur le Web et dans les bases de données.

2.2. Site ARTIST et revues scientifiques

Dès les années 1990, de nombreux projets de publication scientifique ont vu le jour sur Internet, souvent sous l'impulsion de chercheurs impliqués dans les débats sur l'édition scientifique. Parallèlement à ces nouvelles publications et surtout aux nombreux débats qu'elles ont générés, les principaux éditeurs scientifiques ont pris conscience de l'importance du Web pour la diffusion de l'information scientifique. Début 1997, ils ont commencé à mettre en ligne leurs revues. Les acteurs impliqués sont très variés (éditeurs traditionnels, chercheurs, bibliothèques, institutions scientifiques, etc.) et mettent en oeuvre des stratégies différentes. [13]

Lors de la réunion de 11 mai 2005 à Lyon à laquelle j'ai participé et qui concernait le projet ARTIST, des points importants sur la revue électronique dans ARTIST ont été abordés : la revue doit être de haut niveau avec une périodicité (3 à 4 numéros par an), une pagination et un sommaire afin qu'elle est une certaine stabilité de construction ; la présence d'un lectorat est également un critère. Les articles doivent être publiés sur le site, et ensuite sélectionnés pour les mettre dans la revue. Il faut valoriser les auteurs que l'on invite à publier sans oublier de réfléchir à une stratégie internationale de publication.

Les derniers rendez-vous d'ARTIST dans différentes rencontres (CIDE, Ile Rousse, RPIST) ont apporté certaines idées supplémentaires sur la revue dans ARTIST telles que son évaluation et une solution technique en XML/TEI.

Un espace jeune chercheur est ouvert dans le site ARTIST pour les jeunes chercheurs et professionnels de l'IST, leur permettant de déposer les résultats de leurs travaux qui seront ensuite évalués par le comité de lecture. C'est un moyen efficace pour inciter les jeunes chercheurs à alimenter l'observatoire. Au début de mon stage, j'ai publié deux articles dans

l'espace jeune chercheur intitulés respectivement : *Infométrie, Bibliométrie, Scientométrie, un panorama* et *Le projet Européen « E-Biosci »* (Voir annexe 4). Suite à leurs diffusion en ligne, des critiques ou remarques m'ont été faites par des lecteurs externes ou internes à ARTIST sur le fond ou la forme, ce qui m'a permis d'apporter des améliorations à mes articles.

2.3. Site ARTIST et forums

Le site servira de lieu de réflexions pour un ensemble de communautés professionnelles comme le montre la figure 3.

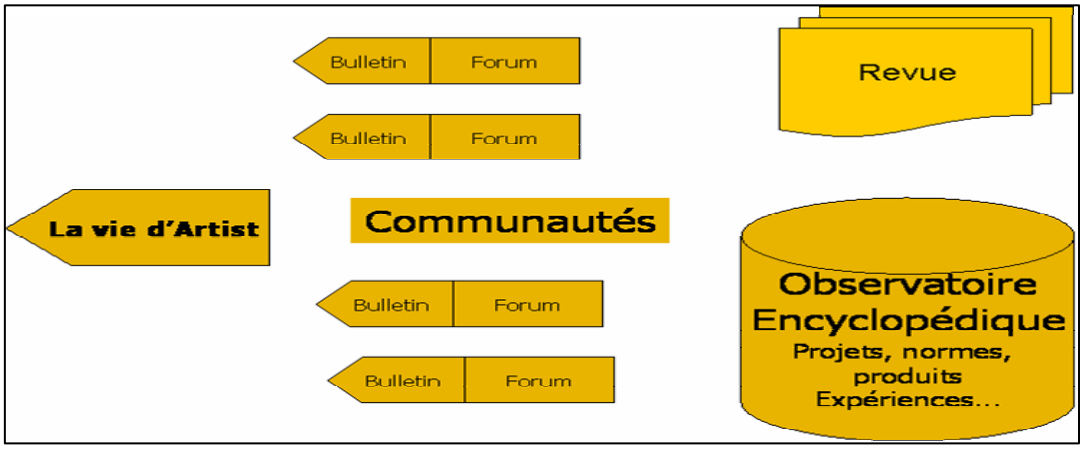


Figure 3 : *représentation schématique du site ARTIST et de son contenu*

Ces communautés échangent des informations sous forme de documents pour s'enrichir mutuellement et manifestent leurs volontés de s'insérer dans le réseau international, comme par exemple celui des archives ouvertes et institutionnelles ou celui des catalogues de bibliothèque.

Dans sa démarche d'appropriation, le site ARTIST s'appuie sur ces communautés spécialisées pour ses activités éditoriales (forum, revue scientifique et observatoire) qui permettront de mettre au service du public des informations liées au transfert de compétences ou de connaissances en IST dans tous les domaines traités dans ARTIST.

- **Organisation des forums**

Les forums représentent un espace qui fait émerger des informations par l'ensemble des articles et des brèves le constituant. Chaque forum aura son bulletin (pour lequel SPIP est bien adapté) périodique permettant de faire régulièrement le point et de donner une

perspective. Tous les acteurs de l'appropriation sont les lecteurs (et rédacteurs) potentiels de l'observatoire comme en autres :

- les chercheurs qui sont au coeur du mécanisme d'appropriation,
- les bibliothécaires et documentalistes de la recherche,
- les chercheurs et ingénieurs en sciences et ingénierie de l'information.

2.4. Site ARTIST et observatoire

Pour permettre une bonne analyse des besoins de la communauté scientifique et un bon repérage de l'état de l'art, un observatoire des nouvelles pratiques de la recherche vis-à-vis des technologies de l'IST a été mis en place dans ARTIST, en se focalisant sur la notion d'« appropriation de la technologie ».

2.4.1. Objectifs de l'observatoire

La partie observatoire permet à un lecteur extérieur d'accéder à des informations classées dans une structure encyclopédique. Il permet également de repérer et de diffuser des expériences intéressantes en matière d'appropriation, d'inventorier des outils et techniques, de donner des informations aux acteurs (chercheurs, bibliothécaires…) sous forme de brèves, de fiches pratiques ou d'articles de synthèse.

2.4.2. Organisation de l'observatoire

L'observatoire sera alimenté par des pratiques de veille avec un fort accent thématique tout en privilégiant la possibilité de coopérer.

- **Thématiques technologiques**

Les thématiques technologiques jouent un rôle fédératif de l'appropriation. On peut y trouver : génie logiciel, infométrie et recherche, métadonnées et architecture, etc.

- **Thématiques scientifiques**

Les thématiques scientifiques constituent un paramètre fondamental de l'appropriation. La structure éditoriale de l'observatoire est organisée suivant les disciplines scientifiques (sciences humaines et sociales, sciences de la vie et médecine, etc).

- **Organisation géographique**

Les pratiques et politiques de recherche sont très diversifiées suivant les régions. On trouve des rubriques organisées dans un cadre géographique, par exemple l'Union Européenne ou l'Algérie dans le monde francophone, etc. [5]

3. Adaptation éditoriale et contribution à la structuration du site ARTIST

La diffusion de l'information sous forme de bulletins, articles, fiches ou brèves, permet une communication directe avec les internautes. La publication en ligne est en effet, un des moyens les plus importants pour l'alimentation de ce site.

ARTIST dispose d'un espace pour publication (espace privé) qui permet aux rédacteurs de diffuser leurs articles en ligne. Cet espace est géré par SPIP, outil de publication collaborative en ligne.

3.1. SPIP outil d'édition collaborative en ligne

3.1.1. Définition

SPIP est un logiciel libre destiné à la conception de site Web, de type système de gestion de contenu. **SPIP** signifie **S**ystème de **P**ublication pour l'**I**nternet. Certains sont tentés de voir dans le dernier p le mot «**P**artagé», dans la mesure où ce logiciel permet surtout d'éditer collectivement un site. [17]

Il permet de gérer le code HTML des pages web (l'architecture de la page, son design, les tableaux, etc) indépendamment de son contenu qui sera stocké dans une base de données. Ce logiciel est écrit sous PHP et utilise la base de donnée MySQL.

3.1.2. Historique

SPIP fut à l'origine créé pour le site http://uzine.net/. Puis les créateurs décidèrent de le livrer sous licence GPL. SPIP s'est peu à peu étoffé ; il est notamment devenu capable de gérer des sites multilingues à partir de la version 1.7. La dernière version stable disponible est la version 1.8.1. [17]

3.1.3. Architecture d'un site SPIP

Un site SPIP est toujours composé de deux parties :

➢ Partie publique

Il s'agit de la partie accessible à tous les internautes. On y trouve toutes les fonctions d'un site internet classique : des rubriques, des articles, un moteur de recherche, parfois des forums, etc. On peut naviguer librement mais il est impossible de modifier le contenu. Voici un exemple de la partie publique ARTIST : http://artist.inist.fr

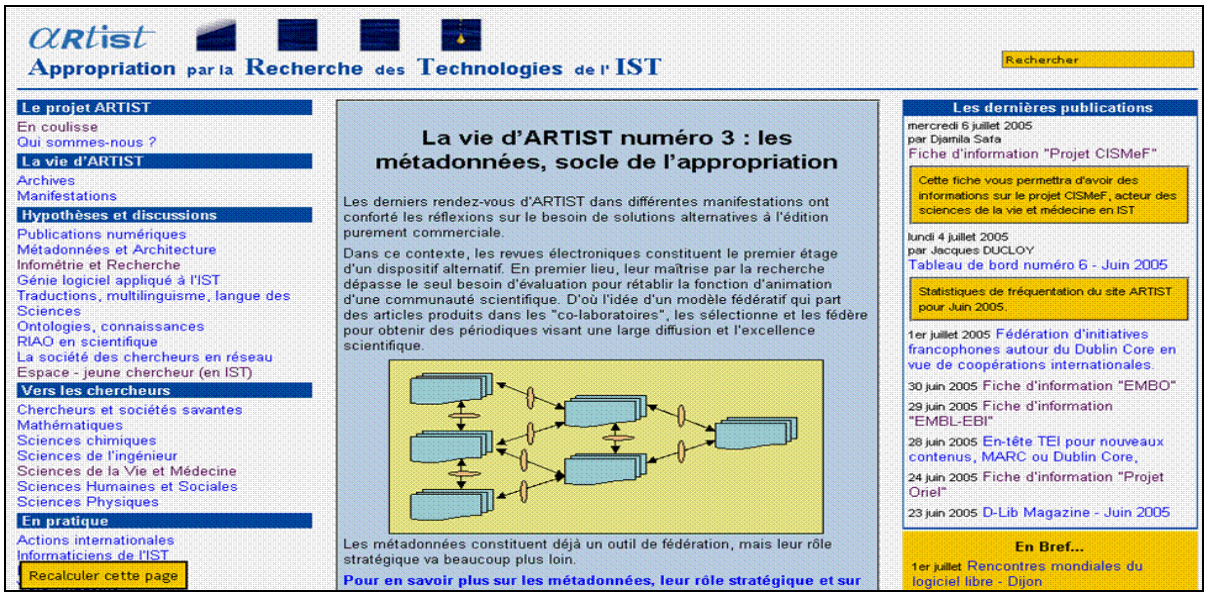


Figure 4 : *Présentation de l'interface du site ARTIST*

➢ Partie privée

Cette partie est destinée à la rédaction et la gestion du contenu. Elle est accessible uniquement aux administrateurs et aux rédacteurs. Pour accéder à la page d'identification, il faut ajouter « /ecrire/ » à l'adresse du site comme par exemple http://artist.inist.fr/ecrire pour ARTIST. Pour accéder à la partie privée de SPIP qui permet de participer à la vie rédactionnelle du site (articles, brèves…), un identifiant et un mot de passe sont attribués par un administrateur du site.

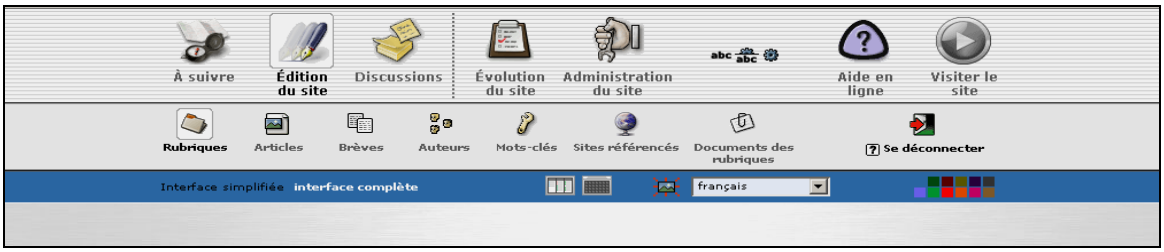


Figure 5 : *L'espace privé dans ARTIST*

3.1.4. Caractéristiques de SPIP

Cet outil permet de gérer un site Web composé principalement d'articles et de brèves insérés dans une arborescence de rubriques imbriquées les unes dans les autres. De plus, il dispense le webmestre et tous les participants à la vie du site d'un certain nombre d'aspects fastidieux de la publication sur le Web.

Pour une gestion éditoriale simple du site par les rédacteurs, des raccourcis typographiques permettent de mettre en forme un texte sans avoir à utiliser le langage ***HTML****, rendant ainsi la contribution rédactionnelle accessible à tous (Voir annexe 5).

En revanche pour le webmestre, l'aspect graphique et la navigation sont définis par des squelettes HTML (ou formats types) définissant par exemple une vue pour la page d'index, une autre montrant une rubrique et un résumé de son contenu, etc.

Pour l'instant on lui connaît très peu d'inconvénients. SPIP est livré avec un format d'interface de navigation complet ; dès qu'on commence à créer le contenu d'un site, il pourra être immédiatement visité, et adoptera l'interface graphique fournie par défaut. [15]

3.2. Conception de fiches d'information

3.2.1. Définition et réalisation

Les fiches d'information créées abordent d'une manière générale un sujet ou un projet dont l'importance concerne les domaines rentrant dans les thématiques traitées dans le site ARTIST.

- **Travail préliminaire**

Dans un premier temps, j'ai travaillé sur la conception d'un modèle de fiches à aspect communicatif qui illustre avec objectivité le thème à traiter. En parallèle, j'ai rassemblé toute la documentation nécessaire concernant le sujet à traiter et plus particulièrement, des projets ayant un rôle majeur en l'IST comme par exemple dans les sciences de la vie et la médecine.

- **Rédaction finale et intégration dans le site ARTIST**
  - **Présentation :** il est préférable, pour la clarté du travail, de préciser les différentes parties de ces fiches : titres du sujet présenté, logos des projets (s'il ya lieu demander l'autorisation avant de les insérer dans les fiches), date de création, auteur, etc).
  - **Contenu :** les différentes parties doivent suivre un plan prédéfini (présentation du projet, objectif, partenaires, références, etc). Toutes ces informations doivent tenir sur au maximum deux pages (Voir annexe 6).

- **Diffusion :** Les fiches ont été mises en ligne après avoir été validées par M. Jacques Ducloy (voir annexe 4).

3.2.2. Difficultés rencontrées

Les difficultés consistent en la présentation et la mise en forme des fiches. Les fiches n'avaient pas le même aspect que celui que j'avais conçu au départ sur Word, du fait qu'une programmation préalable n'a pas été envisagée dans le site ARTIST. Les fiches correspondent à des éléments articles, mais le but c'est qu'elles doivent se différencier des articles réellement publiés sous ce label.

En effet, SPIP propose 3 types d'éléments (rubriques, articles, brèves) mais pas « fiches ». Dans cette partie de travail, j'ai dû intégrer un code HTML lors de la rédaction dans la partie privé du site ARTIST, pour insérer les tables dans mes fiches, comme le montre cet exemple :

➔ Dans l'espace privé du site ARTIST, on insère ce code HTML pour créer les tables

```
<table width="" ALIGN="center"border="3" cellspacing="1" cellpadding="2" COLS="3">
<tr>
   <td bgcolor="00BFFF" width=150><B><CENTER><SMALL>INIST-
CNRS</B></CENTER></SMALL></td>
   <td bgcolor="#C0C0C0" width=300><B><CENTER><img117|center></B></CENTER></td>
   <td bgcolor="00BFFF"width=150><B><CENTER>DATE DE CREATION  :
<SMALL><I>14/06/2005</B></SMALL></I></td>
</tr>
<tr>
   <td bgcolor="00BFFF"width=100><B><CENTER><SMALL>Projet E-BioSci</B></SMALL></td>
   <td bgcolor="#C0C0C0"width=300><B><CENTER>Fiche d'information </CENTER></B>
</td>
   <td bgcolor="00BFFF"width=100><B> <CENTER>Auteur :<SMALL><I>Djamila
SAFA</CENTER></SMALL></I></B>
   </td>
</tr>
</table>
```

Légende : *présentation du code HTML permettant de créer des tables dans les fiches dans la partie privée du SPIP*

➔ Dans l'espace publique du site ARTIST l'affichage est comme ceci :

INIST-CNRS	αRtist	DATE DE CREATION : 14/06/2005
Projet E-BioSci	Fiche d'information	Auteur : Djamila Safa

Légende : *l'affichage de la table construite à partir du code HTML dans l'espace privé du site ARTIST*

Or le but était de pouvoir faciliter l'utilisation de SPIP pour tous les rédacteurs (essentiellement ceux qui n'ont pas de connaissances du langage HTML). L'utilisation de ***CSS**** (feuilles de style en cascade) spécifiquement ne s'appliquant à l'article que si le mot-clé « fiche » lui est attribué, ce mot-clé figure dans le groupe de mots-clés « edition » déjà évoqué.

3.3. Archivage dans une vision encyclopédique

Toutes les informations pérennes (articles ou fiches) sont stockées dans le site dans une vision encyclopédique. Une ventilation par mots-clés est donc utile afin de les retrouver et de faciliter leur consultation.

3.3.1. Indexation

L'indexation est utilisée dans une optique de recherche d'information assistée par ordinateur, sachant qu'une stratégie d'indexation est déterminée par les objectifs fixés en termes de services à rendre aux usagers. L'***indexation**** d'un texte consiste à repérer dans celui-ci certains mots ou expressions particulièrement significatifs (appelés *termes*) dans un contexte donné, et à créer un lien entre ces termes et le texte original. Elle peut être manuelle (faite par un humain), automatique (créée par un programme informatique), ou à divers degrés intermédiaires « assistée » ou semi-automatique (par exemple créée par un humain assisté d'un programme proposant des termes). [16] [18]

Il est possible de créer des ***mots-clés**** liés aux articles ou aux fiches. Par exemple, un article ou une fiche peuvent être liés aux mots-clés «Institutions», « Europe », etc. L'usage des mots-clés permet de proposer une navigation entre différents articles portant sur les

mêmes thèmes. Ainsi, pour que les fiches ou articles soient présents dans différentes rubriques, il a suffit de les indexer par le nom (titre) d'une rubrique ou d'une sous rubrique. Ce fut mon cas lorsque j'ai voulu mettre mes fiches dans la partie « Europe » de la rubrique « Partie encyclopédique ». Ces fiches étaient indexées au départ par le mot-clé « accueil » et n'apparaissent que dans la rubrique « sciences de la vie et médecine » du site.

Note : *accueil : mot-clé invisible sur les fiches qui ne permet l'affichage que dans la page d'accueil dans le site.*

3.3.2. Difficultés rencontrées

La ventilation par mots-clés dans plusieurs rubriques est plus complexe. SPIP classe les articles selon une arborescence classique : un article ne peut appartenir qu'à une seule rubrique. Pour permettre l'affichage automatique d'une fiche dans plusieurs rubriques, nous utilisons là encore les éléments mots-clés SPIP.

A chaque rubrique correspond un mot-clé identique au titre de celle-ci, par exemple : pour un mot-clé « Europe », le nom de la sous rubrique du même nom de la rubrique « Partie Encyclopédique » indexera un article ou une fiche pour qu'ils apparaissent dans la partie « Europe ». Une fiche ou un article, doivent avoir alors, différents mots-clés par rubrique.

Pour contourner ces difficultés, avec la collaboration du webmestre du DPRE, on a effectué quelques essais sous la version SPIP 1.8 (version améliorée de SPIP non encore chargée dans ARTIST, à noter que le site actuel est sous SPIP 1.7).

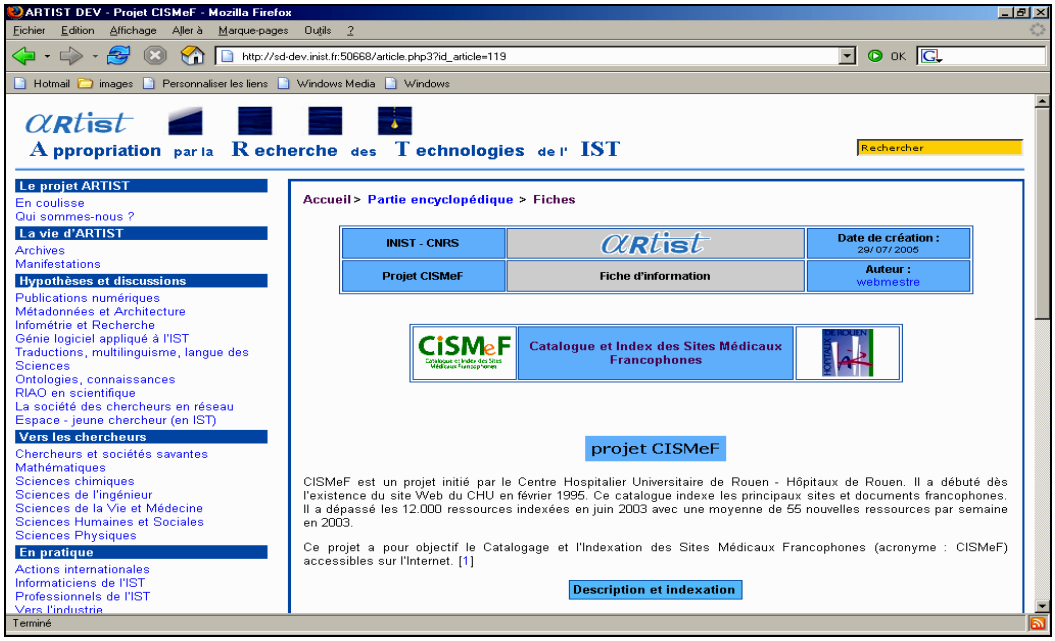
Une programmation spécifique du squelette permet d'appeler les articles ayant le mot-clé correspondant à la rubrique sélectionnée ; par exemple, la fiche X apparaît ainsi à la fois sous la rubrique Y et la rubrique Z, etc. Cette fonctionnalité appelée « contexte de boucle » n'est accessible qu'avec la version SPIP 1.8. La diffusion de ces fiches dans ARTIST a été donc programmée par défaut (Voir annexe 7) ce qui leur donne un aspect identique. (cf. figure 6). Pour permettre une bonne impression de ces fiches, mais également celle des articles, on a dû travailler la programmation en intégrant un code dans le squelette comme le montre cet exemple :

```
/* imprime uniquement  la colonne centrale*/
#entete, #navigation, #colonne2, #recherche, #pied
{
   display: none;
}
#corps
{
   border: none;
}
#colonne
{
margin:0;
}
```

Légende : *code CSS (feuille de style en cascade) intégré dans le squelette et script SPIP*

Tous ces essais ont montré que cette indexation est réalisable. Mais il reste encore la difficulté dans l'emploi des mots-clés contenants un tiret (par exemple : E-BioSci) ou une apostrophe (par exemple : l'Europe). Ces formes ont été évitées.

(A)



(B)

Figure 6 : *Affichage d'une fiche d'information dans le site ARTIST; (A) affichage par défaut,* (B) version imprimable

Remarque : *le navigateur Internet Explorer ne respecte pas complètement les spécifications officielles des normes en usage sur le Web. Les pages conçues dans SPIP peuvent ainsi, parfois, être mal affichées. Certaines pratiques permettent de gérer cette difficulté comme par exemple l'emploi d'une étoile avant l'attribut.*

4. Veille sur le domaine biomédical et l'appropriation des technologies de l'IST

Cette étude de veille va consister à trouver et repérer des références pertinentes sur l'appropriation des technologies de l'IST dans le domaine biomédical afin de mettre de nouveaux articles dans ARTIST.

4.1. Définition de la stratégie de recherche

La Veille Stratégique est le processus par lequel l'organisme observe, recherche, localise, traite, analyse et diffuse l'information stratégique à des fins décisionnelles et compétitives. La particularité d'ARTIST est de suivre la même stratégie de veille mais de rendre public les résultats obtenus de cette veille.

L'objectif de cette étude était de repérer des références significatives, des projets intéressants et innovants en appropriation des technologies de l'IST dans le domaine biomédical.

4.2. Recherche et collecte

Dans une démarche de veille, il est nécessaire d'exploiter un maximum de sources d'informations, afin de ne pas passer à côté des informations importantes.

La recherche d'information est réalisée en utilisant deux ***bases de données bibliographiques**** : PASCAL et ***Pubmed**** (elles disposent de références bibliographiques du domaine biomédical) pour le recueil de notices, ainsi que les moteurs de recherche pour une collecte sur le web, avec une antériorité de 10 ans.

Dans un premier temps, il est fondamental de construire une bonne requête permettant d'obtenir des informations dans le domaine où on effectue notre recherche.

L'interrogation des ***bases de données**** et moteurs de recherche s'effectue par des mots-clés libres ou contrôlés par un thésaurus. Lorsque le terme employé pour la recherche est non contrôlé cela entraîne la collecte de documents non pertinents ("bruit" documentaire), ce fût le cas avec le mot-clé « appropriation ». En revanche, lorsque le terme employé est contrôlé, il faut alors consulter la liste disponible des termes autorisés dans un lexique ou un ***thésaurus****. Par exemple sur le portail ConnectSciences, on peut avoir accès au vocabulaire INIST.

Ce service, basé sur des lexiques terminologiques, nous permet de découvrir le vocabulaire utilisé pour l'indexation des bases de données de l'INIST en Sciences, Technologie et Médecine (PASCAL) et en Sciences Humaines et Sociales (FRANCIS). Il présente les ***descripteurs**** français et le cas échéant leurs équivalents anglais et espagnols ainsi que leurs synonymes.

Cet outil permet aussi de découvrir le plan de classement de PASCAL des informations contenues dans ces bases de données. Le plan de classement est multidisciplinaire, découpé selon les grands domaines scientifiques de ConnectSciences. Il propose également les codes de classement regroupés pour une thématique scientifique donnée. Le code de classement représente un critère de recherche supplémentaire pour interroger les bases de données sur ***serveurs**** et cédéroms. Ces codes de classement sont alphanumériques à l'INIST. Ils sont construits à partir de la hiérarchie qui lie les rubriques aux sous-rubriques (chaque niveau est inclus dans le niveau supérieur).

Pour la base de données Pubmed, le thésaurus ***Mesh**** est utilisé pour l'indexation dans cette base. On peut consulter tous les descripteurs indexés par domaine.

4.2.1. Construire une requête

Pour mes recherches d'information qui consistaient à associer deux concepts, l'information scientifique et technique et le domaine biomédical (sciences biologiques et médicales), j'ai donc procédé comme suit.

- **Requête sur PASCAL :**

Pour construire une requête sur PASCAL, il a fallut que je vérifie les termes utilisés pour l'indexation sur le portail ConnectSciences. Cette vérification m'a permis d'une part de constater que certains termes contrôlés peuvent, ou non, être représentés par un acronyme (IST, NTIC…) et, d'autre part, d'extraire tous les synonymes, par exemple pour « NTIC » on pourra ajouter dans la requête : nouvelle technologie information communication, langage XML, édition électronique, revues numérique, etc. (Voir annexe 8). Cette démarche m'a permis d'optimiser le volume d'information utile à ma recherche.

➢ **Requête sur Pubmed :**

Dans le thésaurus Mesh utilisé pour indexer cette base, j'ai pu extraire la terminologie des domaines qui m'intéressaient (voir annexe 8). Pour montrer l'arborescence de ce thesaurus, voici un exemple :

All MeSH Categories
- Information Science Category
  - **Information Science**
    - Book Collecting
    - Chronology
    - Classification
      - Phylogeny
    - Communication
      - Advertising
      - Answering Services
      - Communication Barriers

Figure 7 : *représentation d'une arborescence de termes Mesh pour le domaine des sciences de l'information. Par exemple, j'ai choisi : communication et information science, etc.*

4.3. Traitement et analyse

4.3.1. Outils et données

L'analyse de l'information suppose le traitement d'un important volume de données qui doit préférentiellement être automatisé. Elle peut inclure des technologies de traitement automatique du langage naturel et de représentation graphique (cartographies) des données bibliographiques.

Pour réaliser cette partie d'analyse, j'ai utilisé différents outils (détaillés ci-après) :

- Miriad pour l'analyse bibliométrique sur le corpus de la base de données PASCAL
- LexiQuest Mine 2.2 ou Leximine, un logiciel propriétaire (édité par SPSS) dédié à l'analyse infométrique pour des données textuelles. Ici on lui proposera deux corpus issus de PASCAL et Pubmed.
- KeyWatch pour l'analyse des URL et la surveillance sur le web.

4.3.2. Procédures d'analyse

- **Miriad**

MIRIAD est une plate-forme bibliométrique qui a pour objet la mise en place d'un système d'aide à la gestion et à la prise de décision. Elle comprend plusieurs modules permettant de faire des statistiques standard sur différents formats de notices, à commencer par ceux utilisés à l'INIST pour les bases PASCAL et FRANCIS. Ces statistiques portent sur :

- les éléments bibliographiques des notices : distribution par code de classement, par type de document, par date de publication, par langue, par affiliation des auteurs et par pays de publication,
- les périodiques : nombre de notices produites par titre et nombre de titres par pays de publication,
- l'indexation : distribution des mots-clés par notices et par fréquence.

- **Outil Leximine**

Leximine est un outil de fouille de données textuelles qui permet l'analyse statistique des documents textuels, le classement et la visualisation sous forme de cartes des termes les plus pertinents dans les corpus PASCAL et Pubmed. De plus, une extraction terminologique est réalisée à l'aide de dictionnaire et de listes que l'on peut modifier de manière à pouvoir extraire les bons concepts et prendre en compte les synonymies et les diverses formes lexicographiques (exemple : génotypique et génotypiques). (cf. figure 8)

Les termes sont classés par type de concept, comme suit :

- **O :** Organization (Organisme)
- **T :** Term (Terme)
- **P :** Person (Personne)
- **U :** Unknown (Inconnu)
- **D :** Product (Produit)
- **L :** location (Localisation)

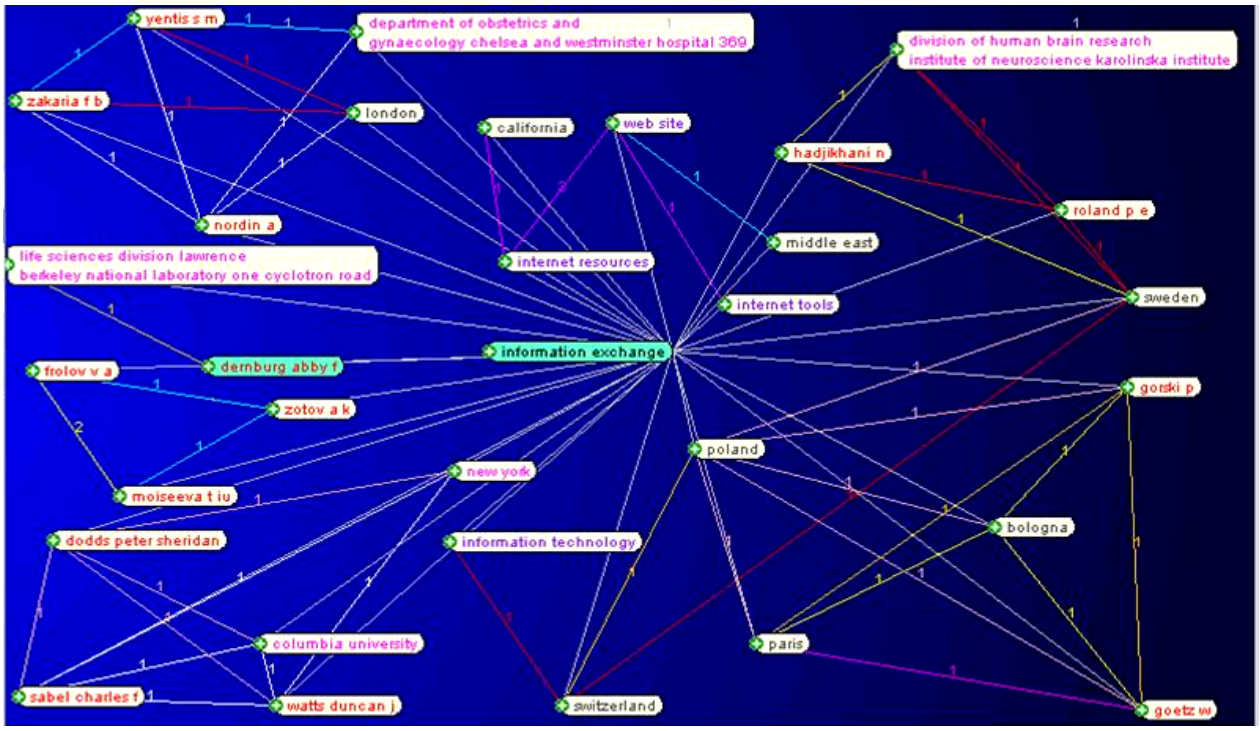


Figure 8 : *interface d'une représentation cartographique sur Leximine. L'interprétation d'une telle carte sera expliquer ci-après.*

- Déchargement des corpus des bases de données vers Leximine:

Les corpus peuvent être déchargés soit en format tabulé serveur (voir annexe 9) soit en format ***SGML**** pour PASCAL et ***XML**** pour Pubmed. La description documentaire est celle fournie par les producteurs des données. Dans ma présente étude, les corpus ont été déchargés sous format serveur pour l'analyse.

Le format d'importation des données sous Leximine est très strict, il a d'abord fallu reformater les données issues de PASCAL et Pubmed.

Leximine ne pouvant singulariser les notices au sein d'un corpus regroupé dans un seul fichier, il est nécessaire de découper le corpus pour avoir chaque notice dans un fichier séparé. Cela se fait en lançant un programme (fournit par Leximine), via l'invite de commande DOS, qu'il suffit de paramétrer (voir annexe 8).

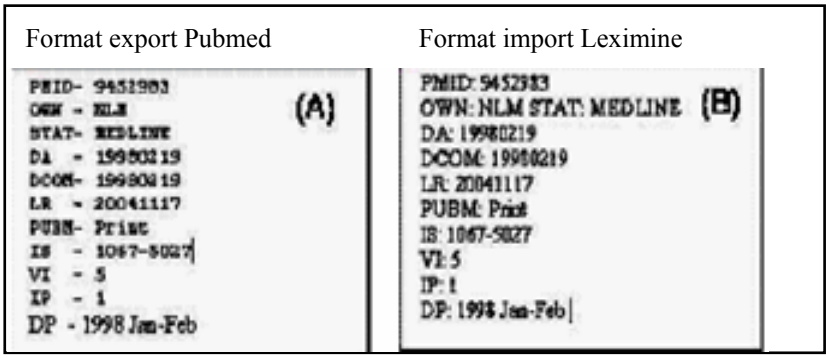


Figure 9 : *traitement des notices avant le découpage du corpus (exemple d'une notice Pubmed) (A) avant traitement, (B) après traitement.*

Une fois le corpus chargé sur Leximine, on y désigne les champs auxquels on s'intéresse par les lettres qui leurs correspondent. Pour cela, il faut aller dans **le module Administration** (cf. figure 10) et sélectionner la liste ***mclef.txt*** de la base Pubmed ou PASCAL. Voici un exemple pour Pubmed :

- **AU**:;P (Person sur le champ Auteur)
- **AB**: //
- **MH**:;C (Concept pour le champ des mots-clés Mesh (indexation propre pour Pubmed)
- **TI**:

Note : *Les champs TI (mots du titre) et AB pour le résumé ne sont pas redirigés vers des catégories de concepts.*

On peut effectuer des modifications dans les dictionnaires intégrés dans Leximine afin de gérer, par exemple, les synonymies, les abréviations des organismes et leurs libellés complets (NCBI pour National Center for Biotechnology Information). Le dictionnaire ***globalcat.txt*** permet de faire des corrections en précisant pour un concept donné sa catégorie : Person (P) pour l'auteur : « colon g a » ou Organization » (O) pour : « CNRS », etc.

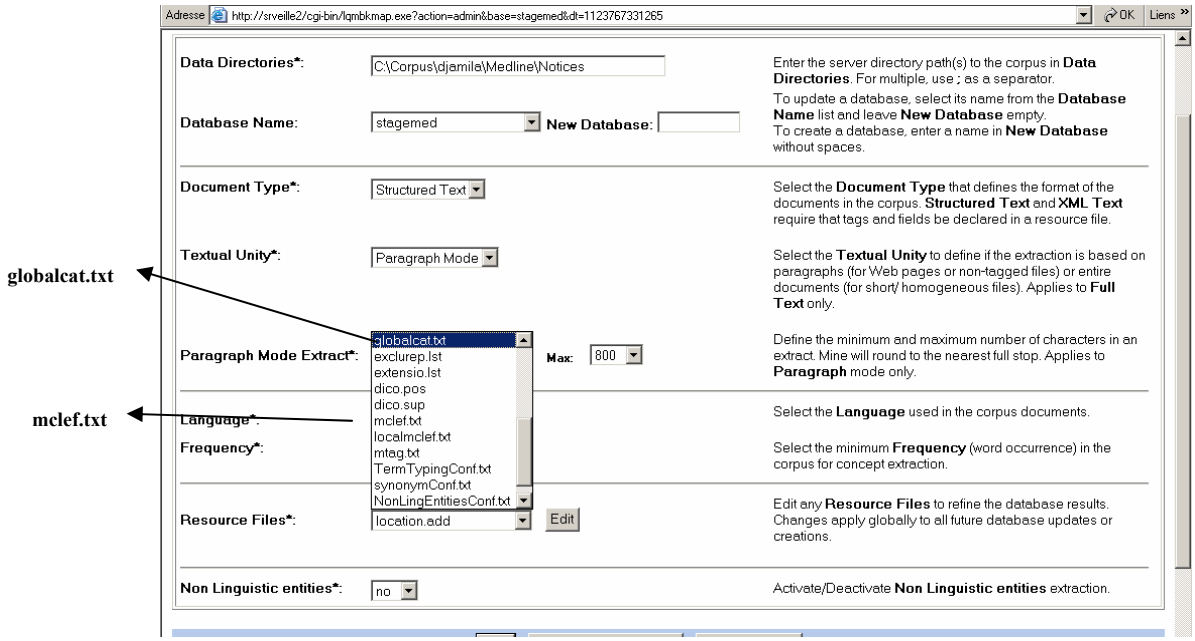


Figure 10 : *Module Administration dans Leximine*

➢ Outil KeyWatch

Pour pouvoir analyser les ***URL**** recueillies sous forme d'un ***bookmark**** (ou signet) à partir de ma recherche sur le web, j'ai utilisé l'outil KeyWatch.

- Plate forme KeyWatch

La plate-forme KeyWatch est un outil de production d'information et d'aide à l'analyse stratégique et concurrentielle. Elle assure le suivi de l'information dans le temps et sa restitution finale sous forme de profils spécifiques. KeyWatch permet ainsi de collecter, d'extraire, de visualiser et de diffuser automatiquement l'information utile (surtout du web).

- Interface de l'utilisateur

Les utilisateurs visualisent les éléments auxquels leurs droits donnent accès et ont chacun un compte propre (base à leurs noms) dans KeyWatch. (cf. figure 11)

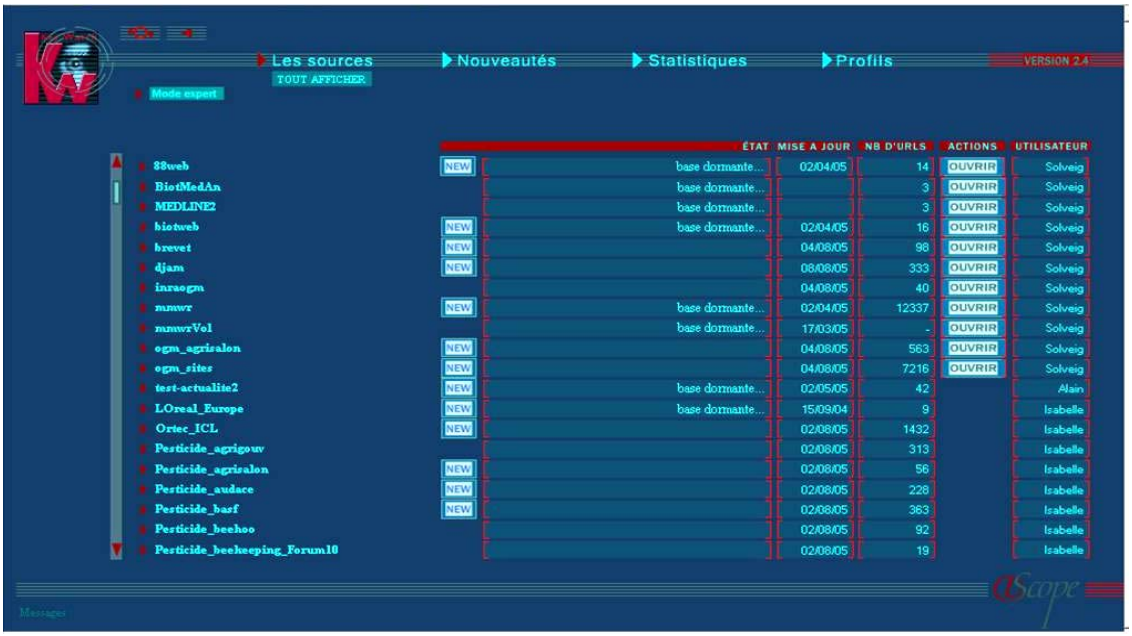


Figure 11 : *interface flash de l'utilisateur sur KeyWatch*

- **Création de la base de données**

Pour charger les URL recueillies pendant la recherche d'informations, des bases d'URL ont été créées sous mon compte. Ces bases permettent de définir les ressources que la plate-forme surveillera. Il est donc nécessaire d'identifier au préalable les sources à surveiller. Le menu « les sources » permet de suivre l'évolution des données à surveiller dans le temps, ainsi que leurs mises à jour (Cf. figure 12).

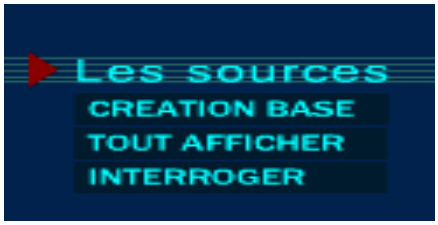


Figure 12 : *présentation du menu « sources » dans KeyWatch.*

- **Interrogation de la base (mots-clés)**

Lorsque les données ont été collectées puis indexées, on peut rechercher au sein de ces bases, grâce à une requête préalablement lancée, des termes ou des expressions à l'aide du moteur de recherche.

- Autres options

On peut également créer des profils qui sont des filtres appliqués sur les nouveautés apparues lors du ***crawl**** de la base d'URL. Ces filtres renvoient une information très ciblée qui est fonction de la présence de mots-clés ou d'expressions dans les documents nouveaux ou modifiés. L'option clustériser peut être activée lorsque le logiciel correspondant est installé (Leximine de SPSS par exemple).

4.4. Résultats d'analyse

Les résultats représentés ci-dessous sont obtenus à partir d'une analyse statistique selon différents critères et d'une analyse infométrique à partir des cartographies sur Leximine.

4.4.1. Résultats obtenus avec Miriad à partir de la base PASCAL

➢ **Distribution des mots-clés par notices**

Rang	Nb. de notices	%	Mot-clé
1	2581	42,291	Base donnée
2	2066	33,852	Homme
3	975	15,976	Internet
4	900	14,747	Système information
5	744	12,191	Europe
6	709	11,617	Informatique biomédicale
7	487	7,980	Article synthèse
8	455	7,455	Information biomédicale
9	424	6,947	Amérique
10	382	6,259	Amérique du Nord

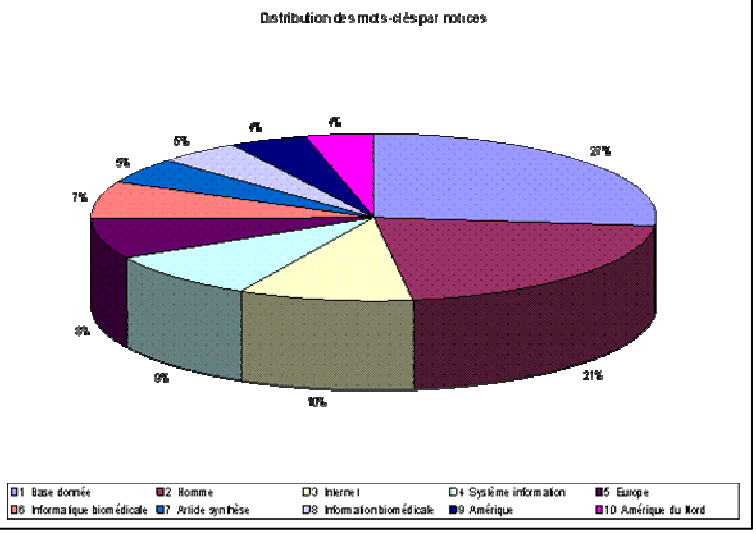


Figure 13 : *distribution des mots-clés dans le corpus étudié (BDD PASCAL)*

L'analyse statistique des termes nous permet de situer leur importance par rapport au nombre de notices les contenant. Or, dans le tableau ci-dessus, on remarque bien la diversité des termes en adéquation avec le sujet traité. Il faut tout de même souligner qu'une exploration à la main est nécessaire pour visualiser les notices indexées par ces termes afin de voir si elles nous intéressent ou pas.

Or dans le graphe ci-dessus (Cf. figure 13), les termes ayant un lien avec le sujet sont : internet, information biomédicale, article synthèse. Pour mieux explorer ces résultats une analyse sur Leximine est utile (ce que je présenterai dans la partie 4.4.2).

- **Distribution des publications par pays de publication (BDD PASCAL)**

1	Pays	Nb. de notices	Pourcentage
2	Etats-Unis	2259	37,2
3	Royaume-Uni	1291	21,25
4	France	889	14,63
5	Allemagne	529	8,71
6	Pays-Bas	443	7,29
7	Reste du monde	299	4,5
8	Irlande	277	4,1
9	Australie	50	0,82
10	Canada	45	0,74
11	Suisse	45	0,74
12	Japon	38	0,62
13	Total	6165	100

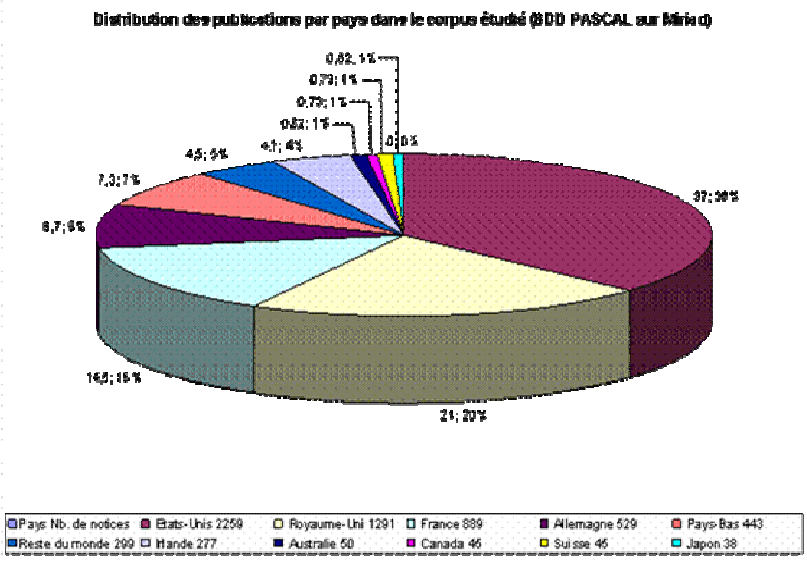


Figure 14 : *distribution pays de publication dans le corpus étudié*

Dans cette analyse si on prend les pays de chaque continent ensemble, on remarque que l'Europe (France, Grande Bretagne…) avec un taux de 56,7% (Cf. figure 14), tient une place importante devant l'Amérique du Nord (USA, Canada, etc) avec un taux de 38% seulement, sans tenir compte des pays du reste du monde (dans lequel on a regroupé tous les pays ayant une fréquence très faible, seuil inférieur à 0,62 %), l'Australie et le Japon avec des taux très faibles (< 1%). Pour ARTIST ces informations sont intéressantes, d'une part pour détecter les pays qui publient sur le sujet étudié dans le monde, d'autre part pour identifier les pays européens (Suisse, Allemagne, Grande-Bretagne, etc) mais également les pays francophones (une partie de Suisse, de Canada, etc) avec lesquels on pourrait envisager une coopération.

4.4.2 Résultats obtenus avec Leximine à partir des bases PASCAL et Pubmed

Les cartographies obtenues sur Leximine m'ont permis d'analyser la thématique étudiée « biomédical et son appropriation des technologies de l'IST » sur les deux bases de données PASCAL et Pubmed. Je me suis donc intéressée en particulier à tous les termes importants ayant un lien avec cette étude.

L'interface graphique permet une visualisation de mots-clés ayant des liens de proximité sémantique plus ou moins forts autour d'un terme d'entrée préalablement sélectionné dans une liste. Cette carte permet, par exemple, de calculer les liens entre les termes et de les associer entre eux. On voit bien qu'entre chaque association il existe un chiffre (Cf. figure 15). Ce chiffre montre le nombre de notices où il y a des ***co-occurrences**** des termes.

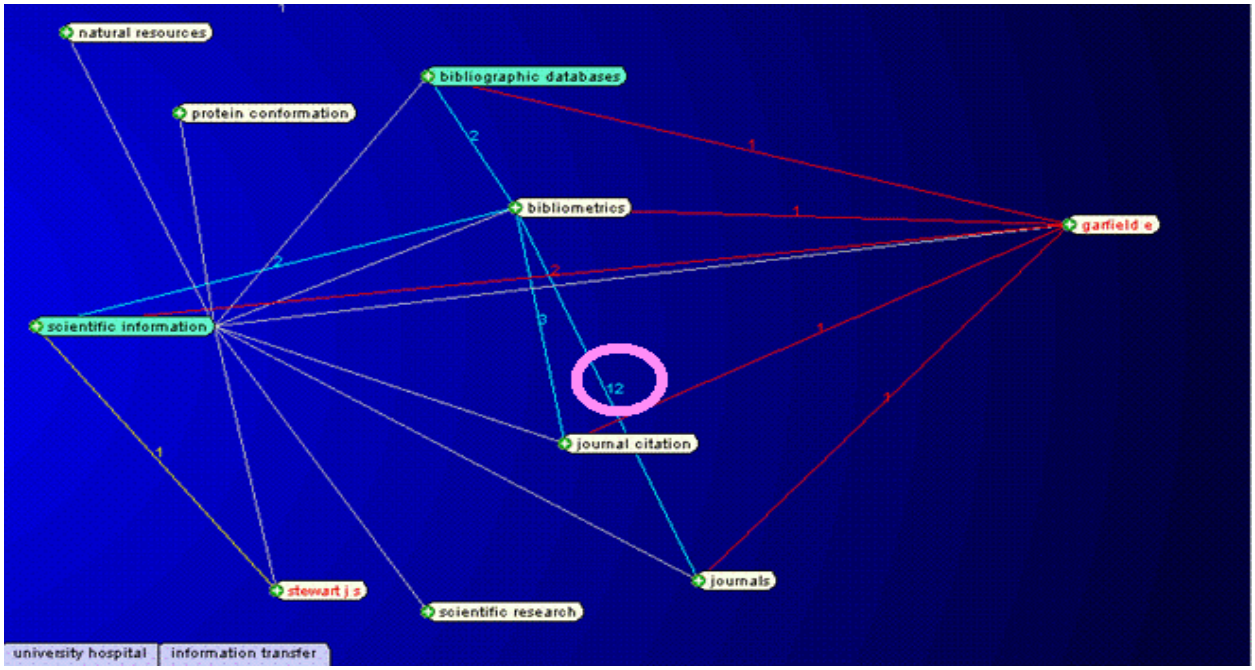


Figure 15 *: carte montrant les associations entre les termes (entre journal et bibliographic databases, le nombre de notices est égale à 12)*

Ce type d'analyse permet également d'identifier des équipes de recherche travaillant sur le domaine analysé, les auteurs qui publient, leurs collaborations, etc. Pour cela, Leximine propose des options (sur la barre d'outils en bas de la carte (Cf. figure 16)) à choisir selon la catégorie dans laquelle on souhaiterait rechercher (Terme, Person, Organisation, etc). Ensuite, on peut réaffiner avec les termes auxquels on s'intéresse. (cf. figure 16)

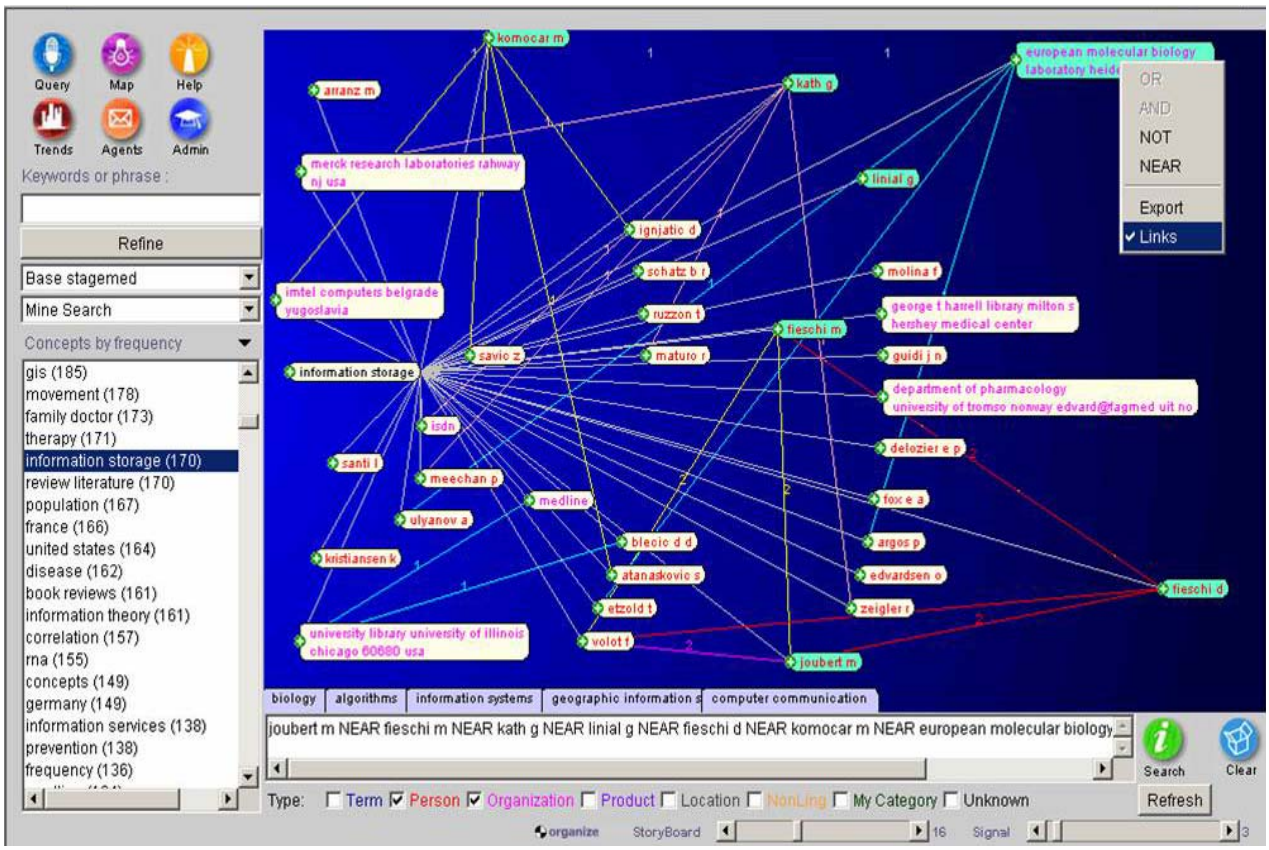
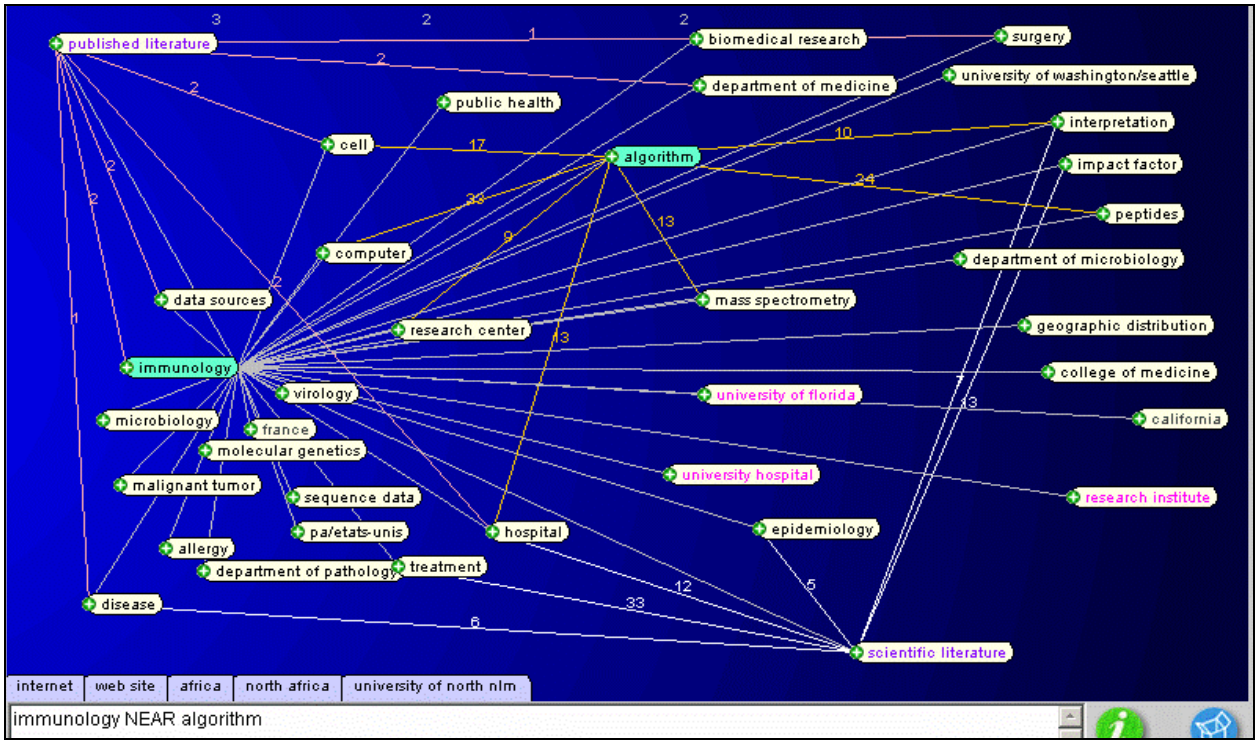


Figure 16 : *interface complète d'une cartographie Leximine*

Sur cette carte, on voit bien les liens (en couleurs différentes) qui existent entre un organisme et un auteur, entre les auteurs, entre les localisations et les auteurs, etc.

Les résultats obtenus à partir de cette analyse vont permettre au demandeur de prendre des décisions dans deux domaines :

- Identifier les principales sous-thématiques du sujet traité et les organismes couvrant tout ou une partie d'entre elles. Ainsi, on va pouvoir s'intéresser à une sous-thématique particulière (exemple immunologie, virologie, microbiologie..) et voir si un organisme ou plusieurs s'y intéressent (cf. Figure 17).



Figue 17 : *carte montrant la sous thématique « immunologie » et tous les termes qui lui sont associés.*

- Repérer les principaux acteurs et collaborations scientifiques travaillant sur le sujet d'étude (cf. figure 18), par exemple ici l'organisme : National Human Genome Research Institute aux USA.

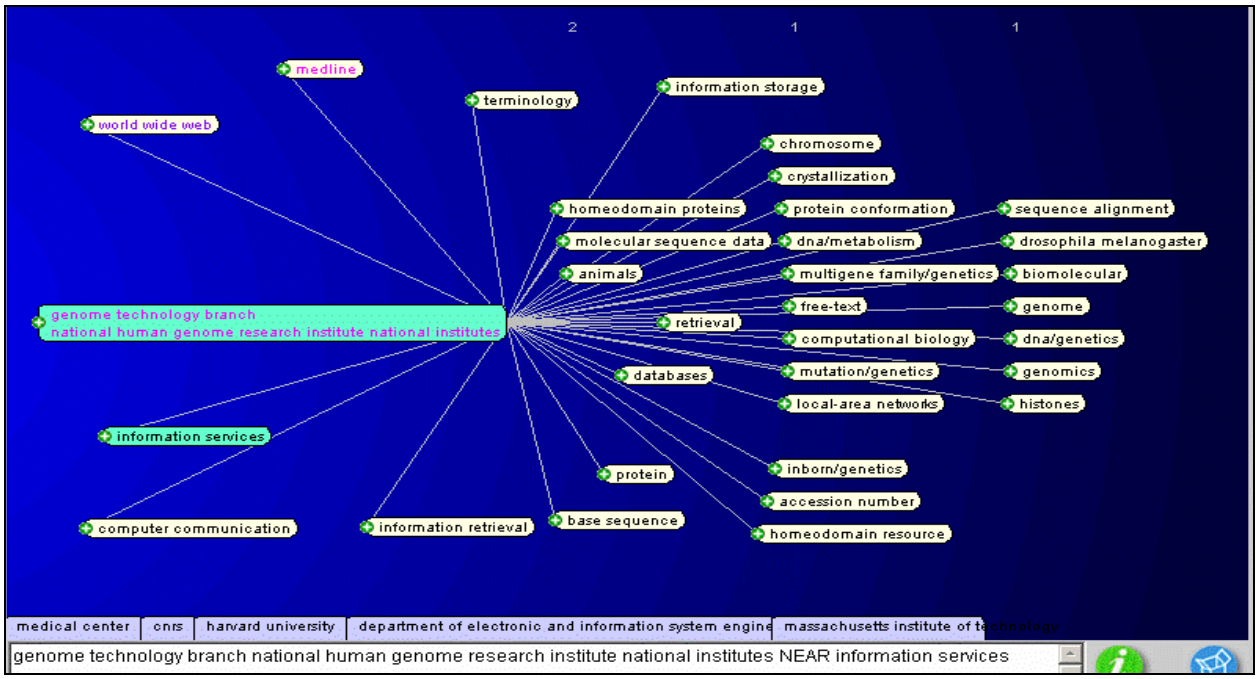


Figure 18 : *carte montrant un organisme travaillant sur le sujet traité*

➢ Base de données PASCAL

- Distribution des termes dans le corpus issu de la BDD PASCAL sur Leximine

Sur Leximine, j'ai pu aller plus loin dans mon analyse d'information, chose que Miriad ne m'avait pas permis (c'est-à-dire une analyse infométrique). J'ai donc choisi de faire une analyse autour du terme « metadata » (métadonnées) ayant un intérêt avec le sujet traité dans mon étude de veille (Cf. figure 19).

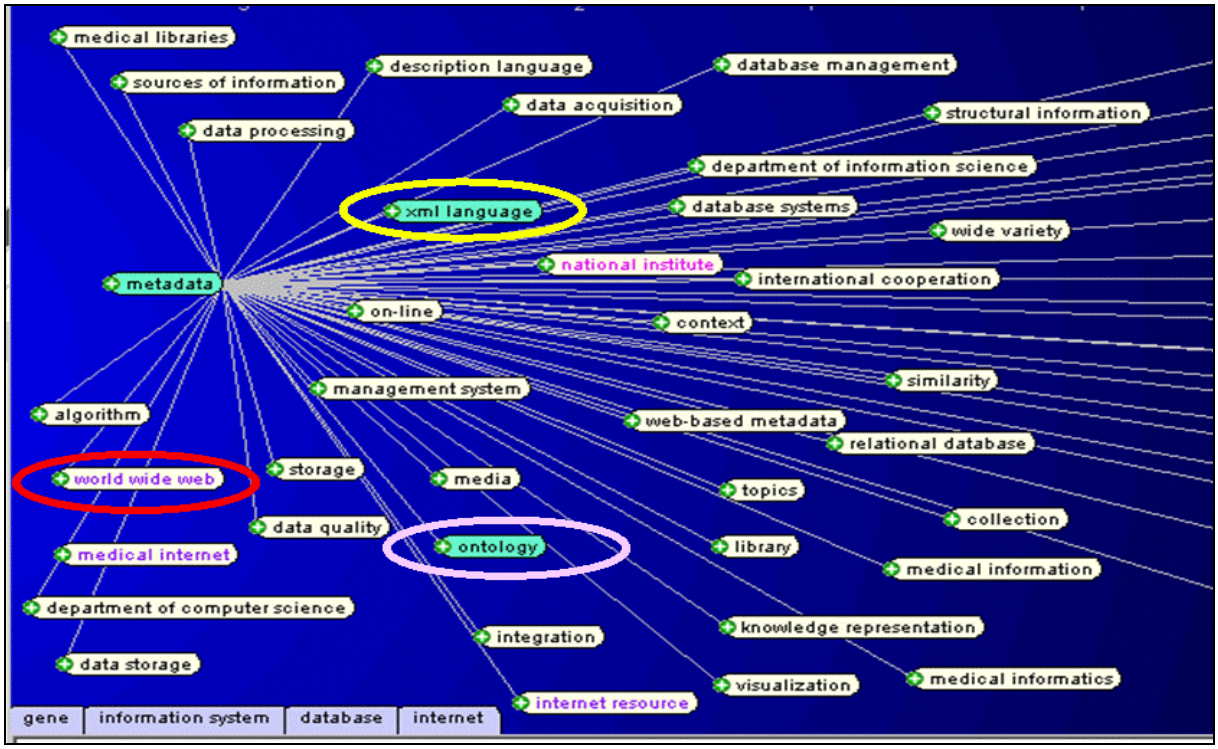


Figure 19 : *cartographies montrant les associations autour du terme « Metadata » issu du corpus étudié (BDD PASCAL) sur Leximine.*

Les résultats de cette analyse sont représentés dans le tableau 1, ici je ne présente que quelques exemples.

Termes	**Quelques références intéressantes**	**Organisme**
XML language	CUELLAR (Autumn A.) et al, 2003 *An overview of CellML 1.1, a biological model description language; Special Issue on Systems Biology and Simulation* http://www.cellml.org/	The Bioengineering Institute University of Auckland Nouvelle-Zélande
World wide web	SCHWEIK (Charles M.) et al, 2005 *The open research system: a web-based metadata and data repository for collaborative research : Information interoperability*	Department of Natural Resources Conservation and the Center for Public Policy and Administration, University of Massachusetts
Ontology	KOHLER (Jacob) et al, 2003 *SEMADA: Ontology based semantic integration of biological databases*	TAG Bioinformatics, Technical Faculty, Bielefeld University; Department of Computer Science, University of Koblenz; AG Cytogenetics/Transcriptome Analysis, IPK-Gatersleben

Tableau 1: *Présentation de quelques résultats d'analyse autour du terme "Metadata" issu du corpus PASCAL sur Leximine*

- Distribution par pays de publication dans le corpus issu de la BDD PASCAL sur Leximine

Les analyses statistiques, représentées par un graphe (Cf. figure 20), sont obtenues à partir de Leximine (colonne de gauche sur la carte représentée dans la figure 21) pour la base de données PASCAL. Sur Leximine, j'ai eu la possibilité de savoir quels termes sont liés au pays et donc leurs relations par rapport au sujet traité dans cette étude (Cf. figure 21).

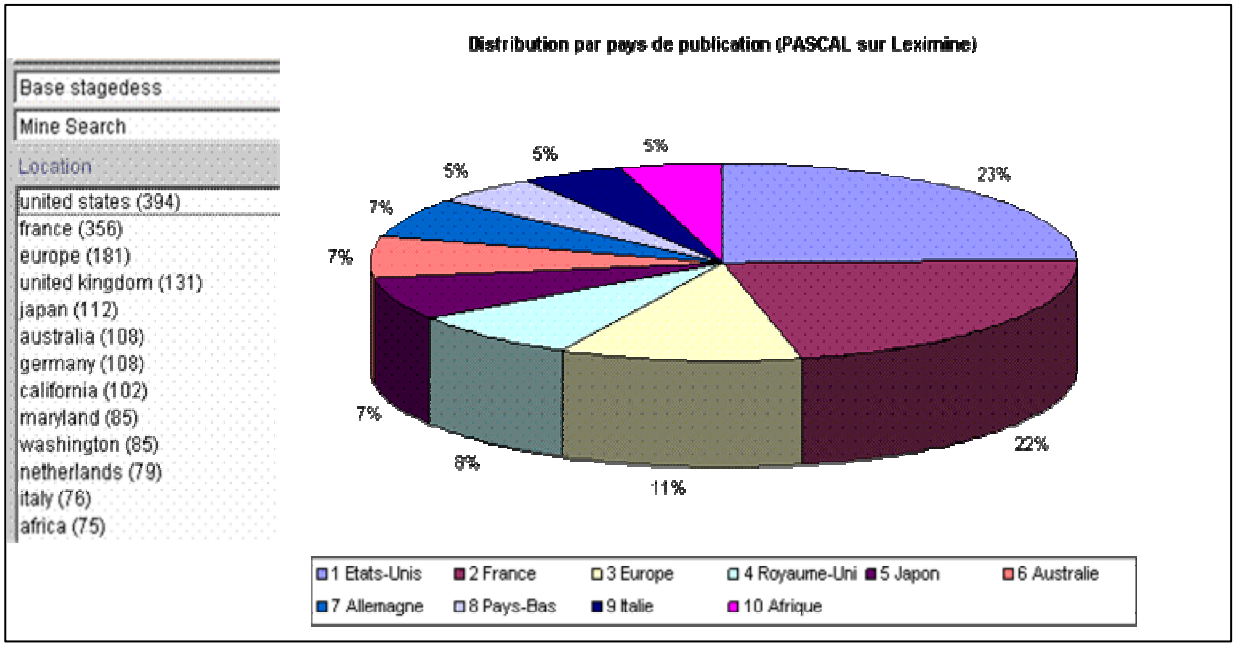


Figure 20 : *distribution par pays de publication (PASCAL sur Leximine) dans le corpus étudié.*

Cette analyse m'a permis de constater que l'Afrique publie aussi dans ce domaine et donc s'intéresse au sujet traité, cette information est capitale car ARTIST s'intéresse non seulement au pays d'Europe mais également aux pays francophones africains. La figure 21 montre bien qu'autour du pays « Afrique », il existe des associations avec des termes liés au sujet traité dans mon étude, en voici un exemple :

Termes	**Quelques références intéressantes**	**Organisme**
Research program	VAN DER KAM (W. J.) et al, 2000 *Effects of electronic communication in general practice*	Department of Medical Informatics, Faculty of Medicine and Health Sciences, Erasmus University Rotterdam
Information dissemination	O'BRIEN (Thomas F.) et al, 2001 *Using internet discussion of antimicrobial susceptibility databases for continuous quality improvement of the testing and management of antimicrobial resistance…*	Department of Medicine and World Health Organization Collaborating Center for Surveillance of Antimicrobial Resistance… USA, Royaum-Unie
Scientifique technical information	ADIGUN (T. A.), 2002 *Analysis of the publications of scientists at the West Africa Rice Development Association, Bouake, Cote d'Ivoire*	West Africa Rice Development Association (WARDA) Côte d'Ivoir

Tableau 2 : *Présentation de quelques résultats d'analyse autour du pays "Afrique" issu du corpus PASCAL sur Leximine*

A partir de ces quelques exemples, on remarque que certaines publications sont publiées par des auteurs africains, cas de l'auteur ADIGUN de côte d'Ivoir.

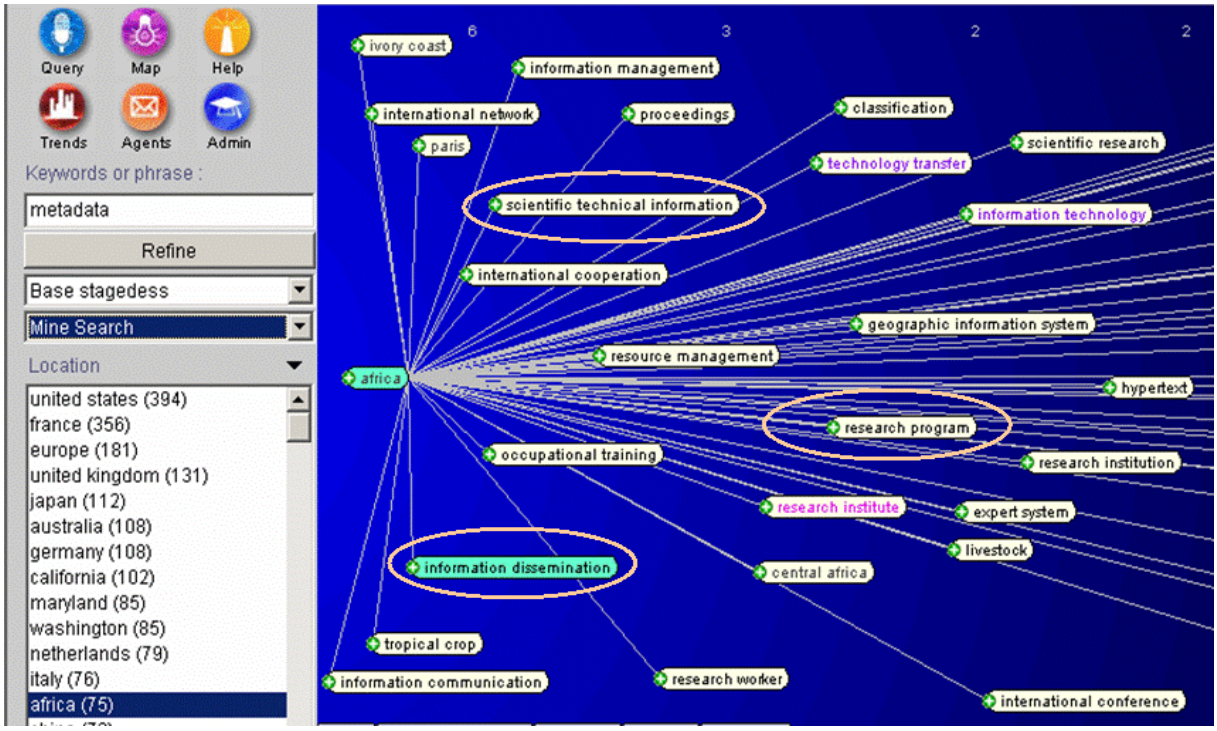


Figure 21 : *Distribution des publications par pays de publication (BDD PASCAL sur Leximine)*

➢ Base de données Pubmed

- Distribution des termes dans le corpus issu de la BDD Pubmed sur Leximine

Comme précédemment pour le corpus issu de PASCAL, Leximine procède à l'extraction terminologique des concepts à partir du corpus de Pubmed et calcule les ***occurrences**** et cooccurrences de ces mêmes concepts.

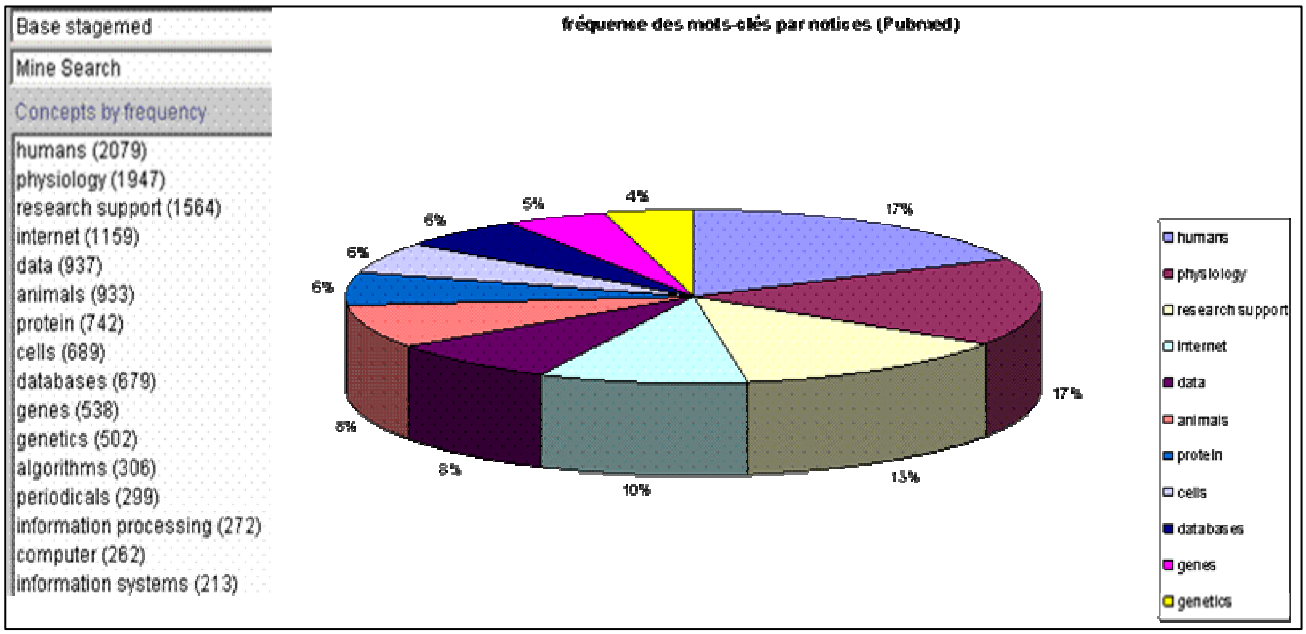


Figure 22 : *Distribution des termes dans le corpus étudié sur Leximine (BDD Pubmed)*

On constate que le terme « Humans » a une fréquence très importante comme le montrent ces résultats, mais je m'intéresse plutôt aux termes ayant une importance directe avec le sujet de mon étude, par exemple : Internet comme le montre la figure 23. Il ne reste plus qu'à explorer les notices manuellement et repérer les plus pertinentes d'entre elles. Cette phase d'analyse m'a permis de retenir certains termes (entourés dans un cadre). Quelques exemples sont représentés dans le tableau 3.

Comme ARTIST s'intéresse aux ressources (francophones ou anglophones) issues des organismes internationaux (monde), j'en conclu que cela pourrait intéresser le groupe ARTIST. On remarque bien que l'Iran et l'Allemagne par exemple, publient sur le sujet traité dans mon étude.

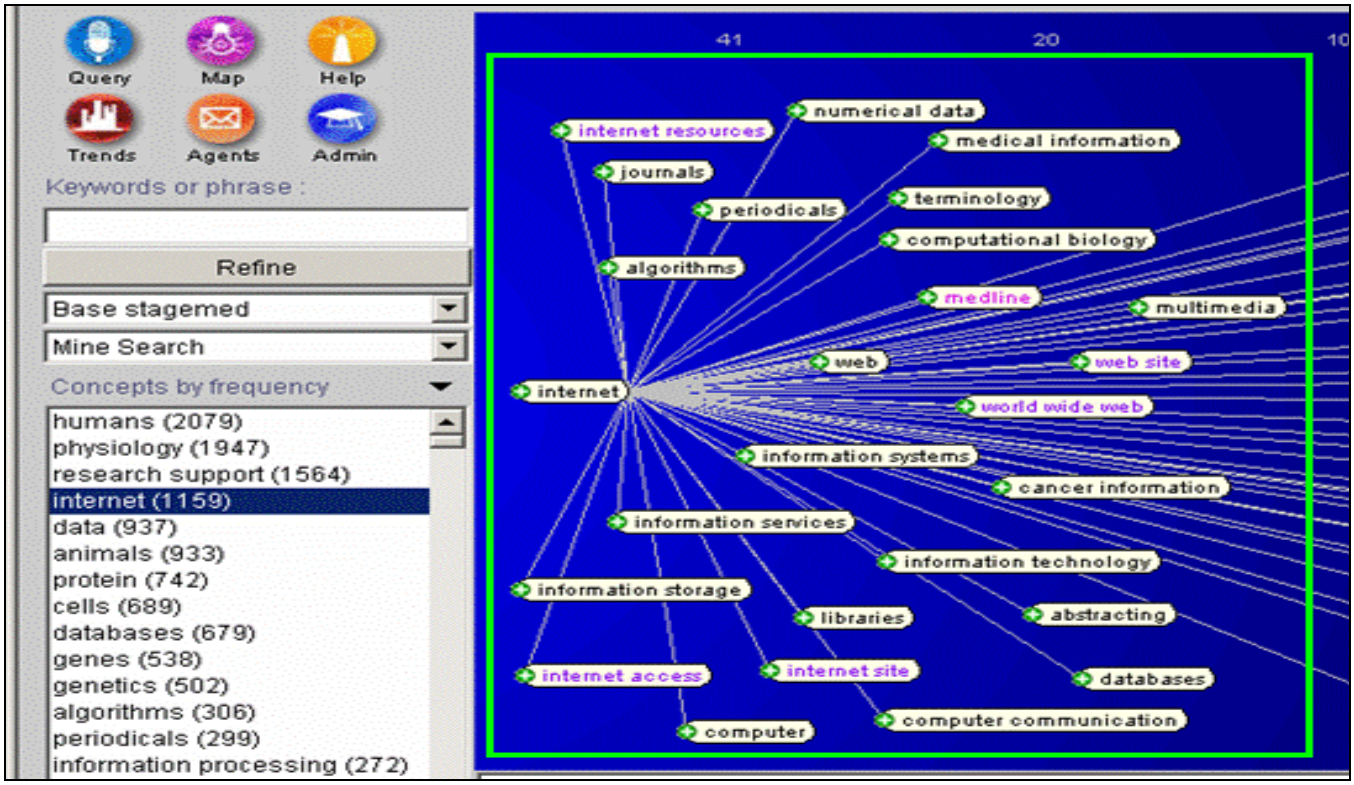


Figure 23 : *carte montrant les associations autour du terme « Internet ».*

Termes	Quelques références intéressantes	Organisme
Information systems	Grune, Barbara and al, 2004 *Report and recommendations of the international workshop "Retrieval approaches for information on alternative methods to animal experiments*".	ZEBET, Federal Institute for Risk Assessment Allemagne
Information storage	Liang, Feng et al, 2004 ORFDB: *an information resource linking scientific content to a high-quality Open Reading Frame (ORF) collection*	Research and Development, Invitrogen Corporation, USA
Journal	Etemadi, Arash et al, 2004 *Views of Iranian medical journal editors on medical research publication*	Endocrine Research Center and the International Journal of Endocrinology and Metabolism, Shaheed Beheshti University of Medical Sciences, Tehran, Iran

Tableau 3 : *Présentation de quelques résultats d'analyse autour du terme "Internet" issu du corpus Pubmed sur Leximine*

➢ Distribution par pays de publication (BDD Pubmed)

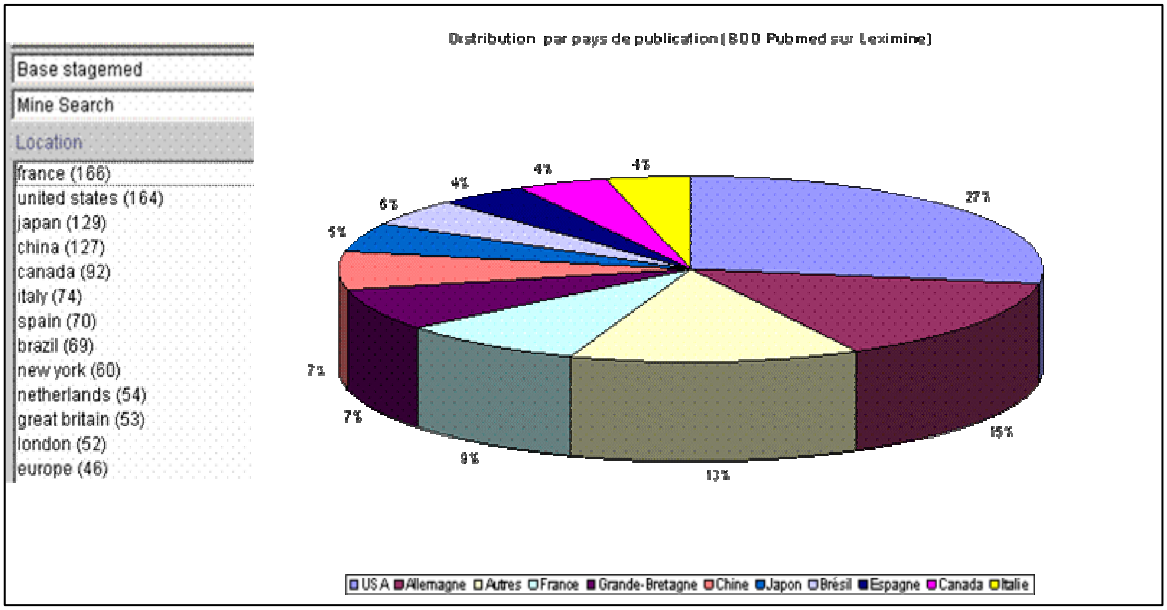


Figure 24 : *Distribution par pays de publication dans le corpus étudié BDD Pubmed (BDD Pubmed sur Leximine)*

Malgré une très forte représentativité de la langue anglaise dans la publication dans Pubmed, la France (Cf. figure 24) est très présente dans l'appropriation des technologies de l'IST dans le domaine biomédical. Suite à l'analyse de la cartographie proposée par Leximine (Cf. figure 25), j'ai choisi d'élaborer avec les concepts suivants :

- France car ce pays est ressorti de l'observation précédente.
- Information language system
- Information storage
- Common access
- Information processing
- Integrated information systems
- Web ressources

Les concepts semblent intéressants dans la mesure où ils ressortent dans la cartographie proposée par Leximine et qu'ils entrent dans le sujet d'étude. Cette première sélection des résultats pourrait intéresser l'équipe projet ARTIST sur plusieurs points :

- Identification d'une équipe de recherche française très présente dans cette problématique. Il s'agit du laboratoire d'immunogénétique moléculaire situé à l'université de Montpellier.

  Au niveau du corpus Pubmed, on observe que la responsable de ce laboratoire a publié 18 fois et co-publié 32 fois. Elle travaille en effet dans les systèmes d'information dans le domaine immunogénétique.
- Repérage de nouveaux mots-clés pouvant indexer les futurs articles dans le site ARTIST.
- Détection de nouveaux projets de recherche sur, par exemple, des méthodes d'indexation des ressources biomédicales utilisant le web sémantique.

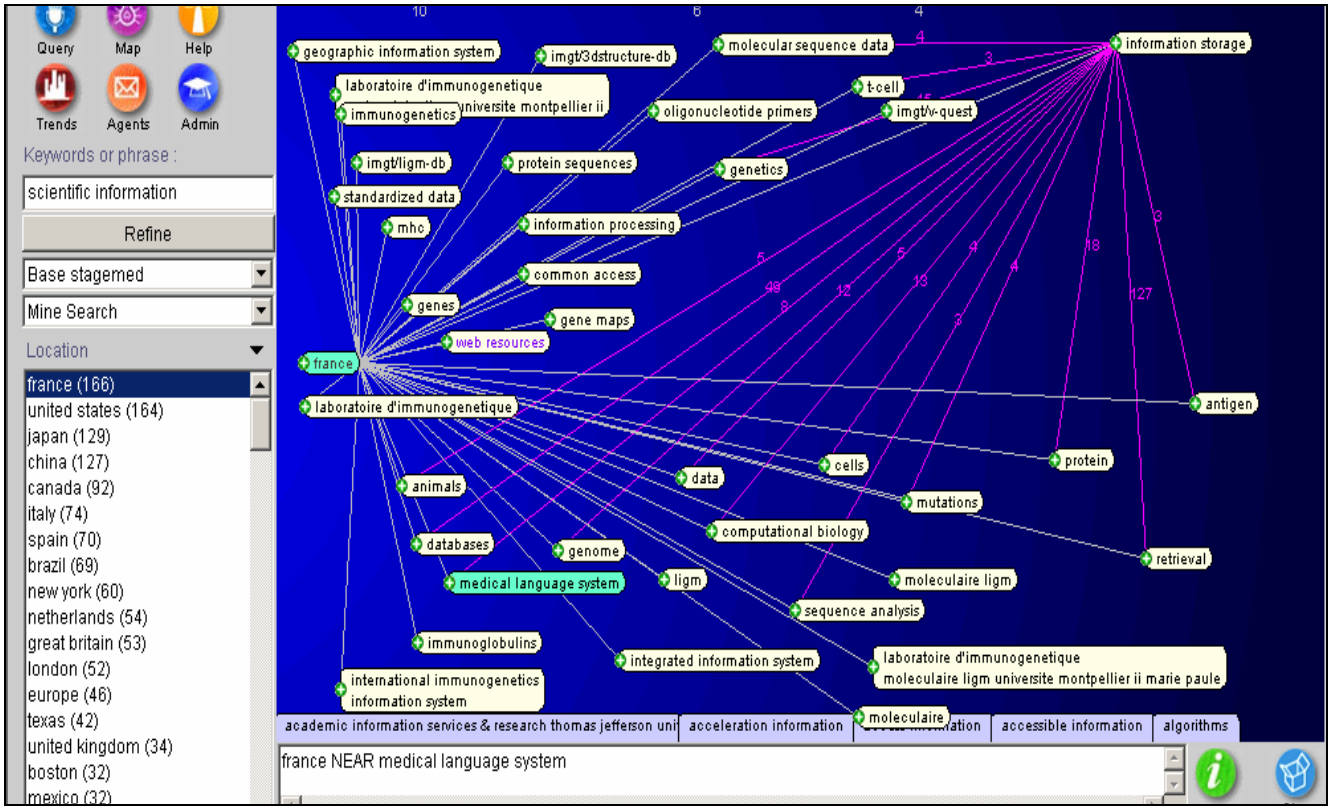


Figure 25 : *Distribution par pays de publication (cartographie Leximine de BDD Pubmed)*
Tous les pays représentant un pourcentage très faible sous le seuil 1 % sont regroupés dans « Reste du monde »

4.4.3. Résultats d'analyse obtenus sur KeyWatch

Pour compléter mon étude de veille, j'ai choisi de faire une recherche sur le web pour identifier les sites et portails concernant l'IST dans le domaine biomédical.

Je m'intéresse à recueillir les nouveautés qui apparaissent sur ce site. Pour cela, il est nécessaire d'utiliser KeyWatch qui est un logiciel de crawl c'est-à-dire qu'il va récupérer toutes les pages du site d'intérêts. De plus, KeyWatch permet de surveiller les modifications des pages existantes, de détecter les nouvelles pages et de signaler les pages supprimées. Pour cela, 340 URL ont été répertoriées dans une base créée sous mon compte dans KeyWatch.

Suite au profil établi sous mon compte avec une requête (Voir annexe 10), Keywatch a signalé de nouvelles URL (Cf. figure 26) intéressante pour mon sujet d'étude. Il s'agit d'une page web « AGROPOLIS-IST » de CIRAD (Centre de coopération internationale en recherche agronomique pour le développement), un des partenaires participant au projet ARTIST. Cette nouveauté m'a permis d'identifier un projet de Plateforme IST Agropolis (PISTA). Cette information peut-être considérée comme une sous-thématique, car ces informations concerne plus le domaine agro-alimentaire que biomédical (thématique traité dans mon étude).

Figure 26 : *A gauche « affichage des URL de la base créée sous mon compte. A droite les nouveautés identifiées des pages web.*

Pour exploiter les pages récupérées en profondeur, il a été nécessaire de faire appel à Leximine, qui est lié à KeyWatch via l'option ***clusteuriser**** de KeyWatch. Cette fois-ci, Leximine va extraire des concepts à partir de documents non structurés (pages web). La figure

27 montre, par exemple, une cartographie proposée par Leximine autour du terme « UMLS » d'entrée, certains concepts ont été repérés et peuvent intéressés l'équipe ARTIST, par exemple :

- mmtx data (initiative d'un programme MetaMap)
- RxNorm (projet Rxnorm qui est un format standard dans le nommage des drogues cliniques)
- semantic network…

J'ai construis une requête, par exemple, « UMLS OR mmtx data » sur un moteur de recherche, je suis tombée sur un site (http://mmtx.nlm.nih.gov/) dans lequel j'ai pu identifiées des informations sur MMTx et on peut y lire les informations suivantes le concernant *: MMTx est une initiative du programme MetaMap disponible pour les chercheurs biomédicaux dans un environnement générique et configurable. MetaMap établit une correspondance entre des textes arbitraires et des concepts dans le métathesaurus UMLS; ou, de la même manière, il découvre des concepts du Metathesaurus présents dans le texte.*

Ces informations peuvent, par exemple, donner lieu à une fiche d'information UMLS à mettre dans la rubrique « Sciences de la vie et Médecine » du site ARTIST ou bien un article de synthèse sur tous les travaux qui tournent autour de UMLS.

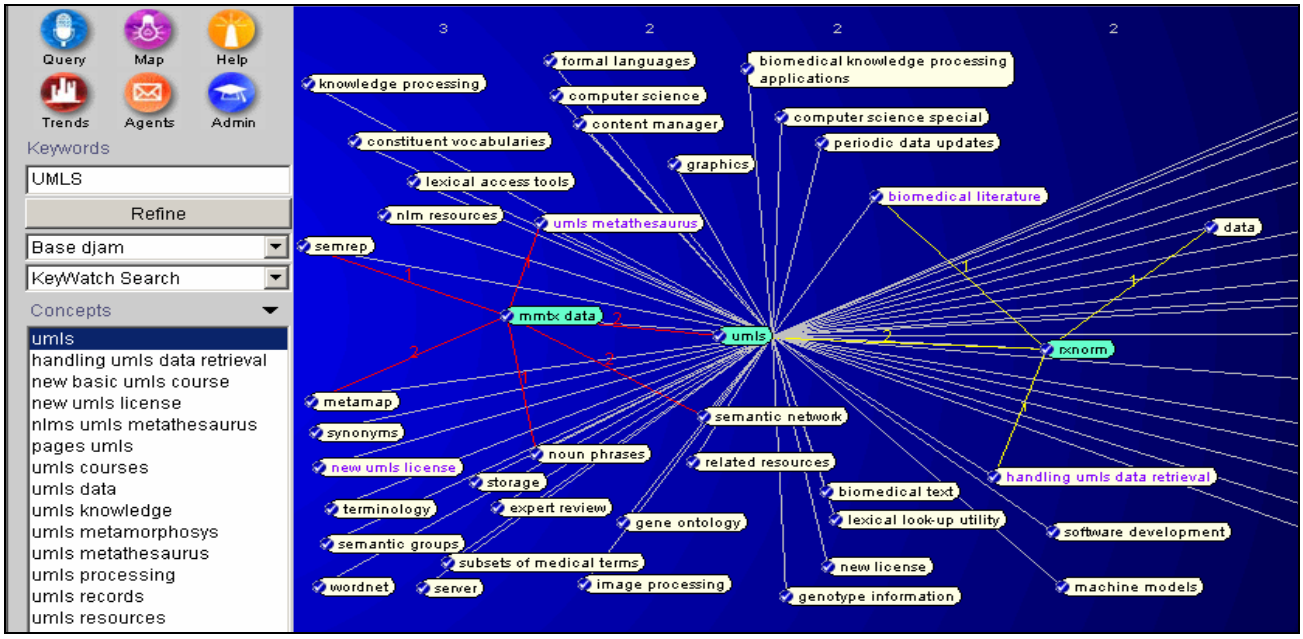


Figure 27 : *carte de visualisation des associations autour du terme « UMLS » sur Leximine de la base créée sur KeyWatch.*

4.4.4. Difficultés rencontrées

- Limites de Miriad

De part sa conception Miriad est un outil permettant seulement une analyse bibliométrique, ce qui ne suffisait pas pour répondre à mon besoin d'analyse. J'ai donc été amenée à utiliser Leximine.

- Limites de Leximine :

La fonctionnalité de catégorisation dans Leximine nécessite souvent des retouches manuelles pour les réajustements. En effet, l'université Claude Bernard considérée initialement comme auteur sera intégrée dans la catégorie « Organisme (O)».

- Limites de KeyWatch :

Les résultats sont lents à obtenir, il faut du temps pour que le logiciel puisse effectuer des mises à jours et aspirer toutes les nouveautés. De plus, il est difficile de le manipuler, l'intervention d'un expert est nécessaire pour effectuer toutes modification des bases de données ou leurs analyse.

- Droits d'accès et disponibilité :

Il m'était difficile d'apporter des modifications sur place sans avoir consulter la personne ayant les droits d'accès, sachant que mes droits d'accès étaient restreints par rapport à l'utilisation des outils de veille (Leximine et KeyWatch). Ce qui m'a pris énormément de temps pour faire cette analyse de veille.

4.5. Diffusion des résultats

Les informations pertinentes repérées au cours de l'étude de veille, donneront lieu à des articles de synthèse, des fiches bibliographique où on recensera toutes les références significatives, d'autres fiches d'information sur des projets ayant une importance capitale dans le sujet d'étude. Toutes ces informations seront diffusées sur le site ARTIST donc elles seront rendues publiques, ce qui constitue une particularité par rapport aux institutions privées.

Voici quelques références classées par date la plus récente

- Bunescu, Razvan et al, 2005
  Comparative experiments on learning information extractors for proteins and their interactions

- SCHWEIK (Charles M.) et al, 2005
  The open research system: a web-based metadata and data repository for collaborative research: Information interoperability and organization for national and global forest information systems.

- Hishiki, Teruyoshi et al, 2004
  Indexing anatomical concepts to OMIM Clinical Synopsis using the UMLS Metathesaurus

- Berman, Jules J, 2004
  Resources for comparing the speed and performance of medical autocoders

- Zou, Qinghua;Chu et al, 2004
  IndexFinder: a method of extracting key concepts from clinical texts for indexing.

Voici quelques projets repérés :

- Projet du génome humain (Human Genome Project)
- GeneNotes- a novel information management software for biologists (BioX program)
- Projet mitochondrie (Mitochondria Project (MITOP))
- Projet du génome *Xylella fastidiosa* (*Xylella fastidiosa* Genome Project)
- Projet du génome Fugu (Fugu Genome Project)
- Projet du génome E. coli (E. coli Genome Project)

Quelques sites intéressants dans le domaine traité :

- http://www.wehi.edu.au/MalDB-www//who.html
- http://lab.digibench.net/transporter/
- http://www.cellbio.com/
- http://genome-www.stanford.edu/cellcycle/
- http://jbiol.com/
- http://www.bentham.org/

- http://imgt.cines.fr/
- http://www.nap.edu/
- http://www.biozone-science.net/

Conclusion et perspectives

Le projet ARTIST lancé début 2005 à l'INIST comporte trois volets : la recherche, la veille technologique et l'édition en ligne sur le thème de l'appropriation par la recherche des technologies de l'IST. Ce nouveau produit ne demande qu'à être amélioré, et alimenté par d'autres informations sur d'autres thématiques que celle traitée lors de mon stage. Les jeunes chercheurs mais également les chercheurs confirmés et les professionnels de l'IST sont invités à venir déposer les résultats de leurs travaux et participer à l'animation du site pour répondre au besoin du public concerné par ce domaine.

Mes missions étaient de contribuer d'une part, à la structuration du site ARTIST en adaptant le squelette et le script dans SPIP et d'autre part, à la gestion du contenu en réalisant cinq fiches d'information de projets, deux articles diffusés dans l'espace jeunes chercheurs et une étude de veille sur l'appropriation des technologies de l'IST dans le domaine biomédical qui m'a permis de repérer des références significatives. Les résultats de ces missions vont permettre l'alimentation du contenu du site par d'autres articles, fiches de projets et des fiches bibliographiques, en particulier pour la rubrique « Sciences de la vie et Médecine », une des thématiques traitée dans ARTIST.

L'absence du personnel pendant l'été a constitué l'une des difficultés dans la réalisation de mes missions, un accès restreint à certains outils de veille. A cela, s'ajoutent des disfonctionnements occasionnels de l'informatique. Le sujet étudié pendant mon stage a constitué une autre difficulté : d'une part les corpus analysés étaient limités à deux bases de données (Pubmed et PASCAL) qui couvrent plus le domaine des sciences de la vie et de la médecine que le domaine des sciences de l'information, d'autre part, le thème de l 'appropriation des technologies de l'IST par la recherche qui est un concept difficile à cerner dans un premier abord.

Ma mission a été prolongée pour une durée de deux mois afin de poursuivre les travaux, d'exploiter plus en profondeur les informations déjà identifiées et d'alimenter le site ARTIST par la rédaction de fiches de projet, articles, etc.

Au final, ce stage m'a permis d'acquérir une bonne pratique des techniques d'édition en ligne avec SPIP et de me familiariser avec les outils de veille tels que Miriad, LexiQuest Mine et KeyWatch. Le travail en collaboration avec le DPRE, le Service veille et les différents ingénieurs de l'INIST m'a donné l'opportunité de travailler en équipe tout en respectant les objectifs de mes missions. De plus, j'ai pu faire preuve de créativité et de rigueur dans la conduite de mon projet. Ma participation, à la manifestation « Rencontre des Professionnels de l'IST » au Palais des Congrès à Nancy et aux réunions de Lyon et Paris dans le cadre du projet ARTIST, m'a beaucoup stimulée et m'a apporté un enrichissement personnel important.

Les porteurs de ce projet prévoient, dans les mois prochains, la mise au point d'une charte graphique, la définition d'un mécanisme d'indexation permettant de simuler un wiki sur une base SPIP, la mise au point de la chaîne éditoriale TEI et le lancement de la revue prévue dans ARTIST.

ABREVIATIONS

- **AEIA :** Atomic Energy International Agency
- **ARTIST :** Appropriation par la Recherche des Technologies de l'IST
- **BDD :** Base De Données
- **CNRS :** Centre National de Recherche Scientifique
- **CSS :** Cascading StyleSheets
- **DESS :** Diplôme d'Etude Supérieur Spécialisé
- **DILIB :** Documentation Information LIBrary
- **DIS :** Direction de l'Information Scientifique
- **DOD :** Departement Of Defense
- **DPRE:** Département Prospective Relations Exterieures
- **HTML :** HyperText Markup Language
- **HTTP :** HyperText Transfer Protocol
- **IMIST :** Institut Marocain d'Information Scientifique et Technique
- **INIST :** INstitut de l'Information Scientifique et Technique
- **IST :** Information Scientifiques et technique
- **MESH :** Medical Subject Heading Section
- **NTIC :** Nouvelles Technologies de l'Information et Communicatio
- **SGML :** Standard Generalized Markup Lanquage
- **SPIP : S**ystème de **p**ublication pour l'**I**nternet «**p**artagé» ou «**p**articipatif »
- **URL :** Uniform Resource Locator
- **VSST :** Veille Stratégique Scientifique et Technologique
- **WWW :** World Wide Web
- **XML:** Extensible Markup Language. Language

GLOSSAIRE

Base de donnée : Ensemble de données organisées en vue de son utilisation par des programmes correspondant à des applications distinctes et de faciliter l'évolution indépendantes des données et des programmes (database en anglais).

Base de donnée bibliographique : c'est une base de type structurée. Elle se présente comme un grand catalogue présentant des revues, articles…Une telle base a pour objet des notices bibliographiques dans lesquelles on a défini les informations (Titre, Date, Auteur…)

Bookmark : Terme anglais, signet(s) : terme français (sur Netscape), favori(s) : terme français (sur Internet Explorer).Utilisé pour sauvegarder des pages web (ou URL) intéressantes.

Clusteurisation (de *Cluster)* : Ensemble de mots ayant des associations fortes, référant à un thème commun.

Crawl : Surveille les sources identifiées suivant l'axe de veille, détecte et signale les nouveautés suivant les critères prédéfinis (tout type de changement, nouveauté correspondante à des mots clés ou requêtes d'alerte, etc.), diffuse les rapports de veille suivant les profils et archive les résultats de veille

Cooccurrence : Présence simultanée de deux ou plusieurs éléments ou classes d'éléments dans le même énoncé.

CSS : Le concept de *feuilles de style* est apparu en 1996 avec la publication par le W3C d'une nouvelle recommandation intitulée *Cascading StyleSheets* (*feuilles de style en cascade*).

Descripteur Mot ou groupe de mots retenu dans un thésaurus (ou vocabulaire contrôlé) et choisi parmi un ensemble de termes équivalents pour représenter sans ambiguïté une notion apparaissant dans un document ou dans une demande de recherche documentaire.

DILIB : *Documentation and Information LIBrary,* plate-forme documentaire développée par l'INIST servant à générer des serveurs d'investigation.

FRANCIS : Base de données de l'INIST. C'est la principale base de données européennes sur les sciences humaines et sociales avec près de 2.5 millions de références bibliographiques depuis 1972.

HTML : *HyperText Markup Language,* il s'agit d'un "langage à balises" (format ASCII), contenant des instructions entre les balises (*tags*) qui sont délimitées entre crochets. Ce langage permet de coder une page à l'aide de commandes de mise en forme. Ces dernières sont ensuite interprétées par un navigateur (*browser*) et apparaissent sur l'écran de l'ordinateur.

http : *HyperText Transfert Protocol*, ce protocole permet à un serveur de communiquer avec un ou plusieurs clients sous la forme de requête et de réponse.

Indexation : Opération qui consiste à décrire et caractériser un document à l'aide d'une analyse des concepts contenus dans ce document.

ISD : *International Science Database* est une base de données bibliographique multilingue et multidisciplinaire couvrant l'essentiel de la littérature mondiale en Science, Technologie, Médecine ainsi qu'en Sciences Humaines et Sociales. Elle est accessible pour les unités CNRS via BiblioSciences, BiblioVie et BiblioSHS.

Medline : C'est la base de données bibliographiques de la National Library of Medicine des Etats Unies. Elle couvre la littérature biomédicale mondiale. Elle contient environ 11 millions de références et couvre une période chronologique qui va de 1966 à nos jours.

Mesh : *Medical Subject Headings* est le thésaurus de la base Medline, base de référence dans le domaine biomédical. Il est produit par la NLM (National Library of Medicine, U.S.A.). C'est à la fois un outil d'indexation, de recherche et de classement.

Mot-clé : Mot ou groupe de mots choisi : soit dans le titre ou le texte d'un document, soit dans une demande de recherche documentaire, pour caractériser le contenu.

Occurrence : L'occurrence d'un mot est égale au nombre de fois où un mot apparaît dans un texte.

PASCAL : Base de données multidisciplinaires et multilingue de l'INIST. Cette base de données signale près de 14 millions de références bibliographiques en sciences, technologie et médecine depuis 1973.

Pubmed : C'est une interface développée par le NCBI (National Center for Biotechnologie Information) de la NLM et les éditeurs de revues biomédicales. Elle permet d'accéder gratuitement aux bases Medline et Premedline.

RTP-DOC : *Documents et contenu : création, indexation, navigation,* c'est une plateforme d'échange participative utilisée par les collaborateurs et les laboratoires impliqués dans l'activité du réseau, ou simplement intéressés par la thématique du document numérique. Un réseau des départements STIC et SHS du CNRS.

Serveur : Ordinateur qui met ses ressources à la disposition d'autres ordinateurs sous la forme de services, qui peuvent être : Espace disque, Information, Base de données, Traitements automatisés.

SGML *: Standard Generalized Markup Language*, c'est un langage structuré normalisé international pour la documentation technique (ISO 9979 : 1986) qui permet de décrire la structure logique d'un document.

STANALYST : Plate forme logicielle regroupant différents outils (serveur Miriad, la plate forme ILC, les logiciels NEURODOC et SDOC et le système HENOCH)

TEI : *Text Encoding Initiative*, c'est une initiative qui, à l'origine, visait à décrire, avec une structurartion SGML/XML, les textes existants.

Thésaurus : Vocabulaire contrôlé et dynamique de termes ayant entre eux des relations sémantiques et génériques, et qui s'applique à un domaine particulier de la connaissance.

URL : *Uniform Ressource Location*, c'est un format de nommage universel pour désigner une ressource sur internet.

Web : Abréviation de *World Wide Web*, c'est un ensemble site informatique de type multimédia interconnecté pour constituer une hyper-base de données mondiale accessible en utilisant un navigateur (Netscape Communicator, Internet Explorer).

XML : *Extensible Markup Language. Language* de balisage établi pour répondre au besoin d'élargir la nature des documents à échanger sur le Web, de faciliter l'interopérabilité entre applications, et de permettre des descriptions plus précises.

Références bibliographiques

1. **AYMONIN David, 1996**
   « L'Information scientifique et technique (IST) sur Internet. Ressources disponibles et méthodes de recherche »
   Article original paru dans l'ouvrage « *Les Nouvelles Technologies dans les bibliothèques* » sous la direction de Michèle ROUHET - Ed. du Cercle de la Librairie ; 1996. - ISBN : 2-7654-0623-5".
   http://www-scd-ulp.u-strasbg.fr/urfist/IST_Internet/ist_plan.htm
   Site consulté le 30 juillet 2005

2. **Cefrio, 2005**
   « Les technologies de l'information et de communication doivent être rentables non seulement économiquement mais surtout humainement »
   Onzième assemblée annuelle du CEFRIO
   http://www.cefrio.qc.ca/communiques/commun_17_imp.cfm
   Site consulté le 29 juin 2005

3. **ConnectScienses-INIST/CNRS, 2005**
   http://connectsciences.inist.fr/services/divers/aid_about.php
   Site consulté le 22 juin 2005

4. **Ducloy Jacques, 2005**
   « ARTIST : Proposition » (en ligne : 31 janvier 2005)
   http://artist.inist.fr/article.php3?id_article=1 (Site consulté le 25 mai 2005)

5. **Ducloy Jacques, 2005**
   « Les axes de l'observatoire » (en ligne : 4 février 2005)
   http://www.artist.inist/articlefond.php3?id_article=11 (Site consulté le 22 mai 2005)

6. **Ducloy Jacques, 2005**
   « Démarrage de l'initiative ARTIST (Appropriation par la Recherche des Technologies de l'IST) » (en ligne : 21 février 2005)
   http://rtp-doc.enssib.fr/article.php3?id_article=186 (Site consulté le 22 mai 2005)

7. **Giffard Alain**
   « Logique de l'appropriation »
   http://alaingiffard.blogs.com/culture/2005/04/logique_de_lapp.html
   (Site consulté le 20 juillet 2005).

8. **INIST**
   http://www.inist.fr (Site consulté le 25 mai 2005)

9. INIST, 2005
SPECIAL'IST-A propos de l'INIST
http://www.inist.fr/actu/inist_cest.php (Site consulté le 25 mai 2005)

10. INIST, 2004
« L'information pour un monde de recherche»
Brochure publiée en juin 2004

11. INIST, 2003
« Institut de l'Information Scientifique et Technique »
Brochure publiée en 2003

12. Panijel Claire, 1999
« Information scientifique et technique »
URFIST de Paris/ecole des Chartes
http://www.ccr.jussieu.fr/urfist/def-ist.htm (Site consulté le 27 mai 2005)

13. PIGNARD Nathalie, 2005
« La publication scientifique sur Internet »
http://archivesic.ccsd.cnrs.fr/documents/archives0/00/00/14/14/sic_00001414_02/sic_00001414.html (site consulté le 29 juin 2005)

14. SG Hebdo, 2005
« La publication du secrétariat général du CNRS aux unités de recherche et de service » 8 juillet 2005- n°71
http://www.sg.cnrs.fr/sghebdo/archives/2005/07/080705/sghebdo.htm
Site consulté le 18 juillet 2005

15. SPIP
« Documentation SPIP »
http://www.spip.net/fr_article464.html (Site consulté le 29 mai 05)

16. URFIST
«L'indexation automatique »
http://www.urfist.cict.fr/indexation.html (Site consulté le 29 mai 05)

17. Wikipedia, 2005
«Système de publication pour l'internet partagé»
http://fr.wikipedia.org/wiki/SPIP (Site consulté le 29 mai 05)

18. Wikipedia, 2005
«Indexation»
http://fr.wikipedia.org/wiki/Indexation (Site consulté le 29 mai 05)

ANNEXES

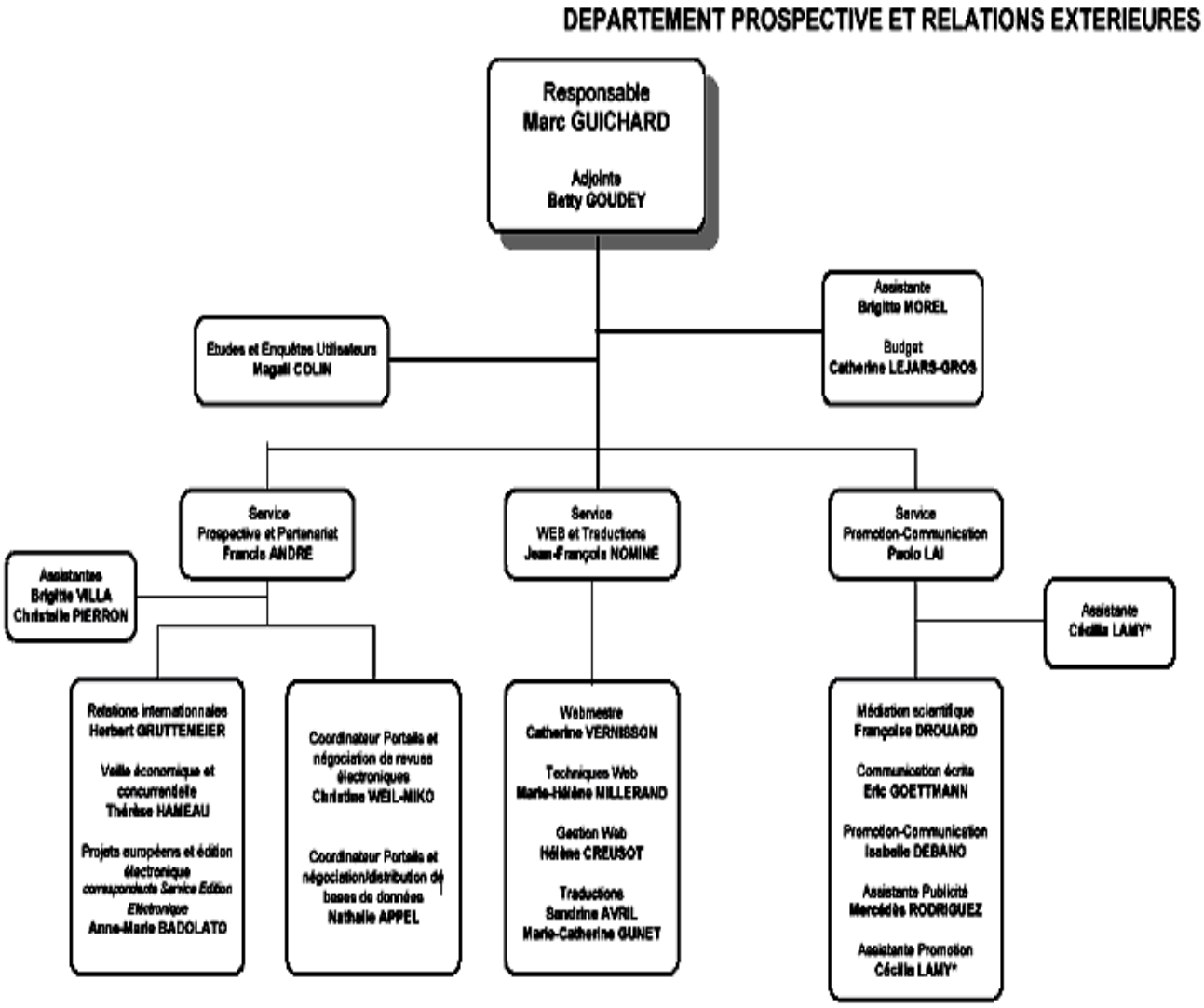
Annexe 1

Les adresses utiles pour les services INIST:

- http://services.inist.fr/
- http://connectsciences.inist.fr/
- http://www.inist.fr/bibliosciences/
- http://thesa.inist.fr/
- http://articlesciences.inist.fr/
- http://bibliovie.inist.fr/
- http://publicnrs.inist.fr/
- http://www.captaindoc.com/
- http://biblioinserm.inist.fr/
- http://titanesciences.inist.fr/
- http://biblioshs.inist.fr/

Annexe 2

Organisation de DPRE (Département Prospective et Relations Extérieures)



Remarque :

Attention l'organigramme de l'organisation de DPRE à l'INIST-CNRS va évoluer d'ici janvier

Annexe 3

Comité scientifique et de pilotage

Un premier noyau de comité de pilotage a été mis en place, en voici les représentants de chaque organisme qui y participe :

- Jacques Ducloy, INIST, porteur du projet

Un noyau fondateur :

- Sylvie Lainé-Cruzel, Professeur en sciences de l'information et de la communication ERSICOM-Lyon 3
- Francis André, responsable du service prospective et relations internationales de l'INIST

Des représentants de formations partenaires :

- Lucile Grasset, CIRAD
- Ghislaine Charttron, ENSSIB
- Amos David, responsable de l'équipe SITE du LORIA

Des représentants de communautés scientifiques à titre personnel et/ou institutionnel :

- Bernard Dousset-IRIT, représentatnt de la communauté VSST
- Eric Trupin, université de Rouen, représentant de la communauté CIDE
- Jean-Paul Ducasse, université de Lyon 2

Des membres de la communauté internationale :

- Christine Vanoirbeek, EPFL (Suisse)
- Jean-François Nominé, webmestre de l'INIST, animateur de l'équipe technique d'ARTIST.

Annexe 4

Liste des articles et fiches d'information parus sur le site ARTIST :

- Articles

- ***Le projet Européen « E-BioSci »***
  http://artist.inist.fr/article.php3?id_article=118

- ***2. Infométrie, Bibliométrie, Scientométrie, un panorama.***
  http://artist.inist.fr/article.php3?id_article=105

- Fiches

- ***Fiche d'information "Projet E-BioSci"***
  http://artist.inist.fr/article.php3?id_article=124

- ***Fiche d'information "Projet Oriel"***
  http://artist.inist.fr/article.php3?id_article=126

- ***Fiche d'information "EMBL-EBI"***
  http://artist.inist.fr/article.php3?id_article=135

- ***Fiche d'information "EMBO"***
  http://artist.inist.fr/article.php3?id_article=138

- ***Fiche d'information "Projet CISMeF"***
  http://artist.inist.fr/article.php3?id_article=159

Annexe 5

Mémento pour l'enrichissement typographique dans SPIP (http://www.spip.net/IMG/pdf/doc-273.pdf)

Mémento pour l'enrichissement typographique dans SPIP

Ce tableau rappelle brièvement les méthodes à utiliser pendant l'écriture d'un article pour ajouter des titres, du texte en gras, en italique, etc..., sans avoir à connaître HTML. L'utilisation d'HTML reste néanmoins possible, et nécessaire pour les cas plus complexes.

Fonctionnalités	***Méthodes***	***Commentaires***
Intertitre	{{{ le titre }}}	Le texte entre triples accolades est affiché comme un titre.
Changement de paragraphe	Passer une ligne	
Caractères gras	{{texte en gras}}	Le texte entre double accolades apparaitra en gras.
Caractères en italique	{texte en italiques}	Le texte entre simples accolades est affiché en italique. Astuce: pour avoir du texte en gras et en italique, mettre trois accolades en insérant une espace (pour faire 1+2), et en respectant la symétrie. Ex : { {{ texte en gras et en italique }} }
Liste à puces	- premier élément - deuxième élement - etc...	Le petit trait sera automatiquement remplacé par une puce, telles qu'elle est définie dans la charte graphique.
Trait de séparation	----	Entrer une ligne contenant quatre petits tirets (au moins).
Lien hypertexte	[texte -> URL]	Mettre le texte et l'URL entre crochets, séparés par une flèche. Le texte devient une zone cliquable, et renverra le lecteur sur la page web indiquée par « URL ». Ex : [le site national de l'AITF->http://www.aivf.asso.fr]
Lien hypertexte vers un article	[texte->n°d'article]	Pour renvoyer vers un article du site, il suffit de donner le numéro de l'article. Par exemple, en supposant que l'article 12 donne la liste des contacts : [nous conctacter->12]
Lien vers une rubrique	[texte->rub*xxx*]	Où xxx est le numéro de la rubrique. Ex : [voyez notre agenda->rub3]
Lien vers une brève	[texte->br*xxx*]	Où xxx est le numéro de la brève. Ex : [Annonce de l'AG->br25]
Note de bas de page automatique	[[texte de la note]]	Le texte entre doubles crochets droits apparaîtra en bas de la page, et sera remplacé par un numéro généré automatiquement.
Note de bas de page numérotée	[[<x>texte de la note]]	Le texte entre doubles crochets apparaîtra en bas de page, et sera remplacé par le numéro *x* indiqué entre « < > ».
Faire un tableau simple	| aaa | bbb | ccc | | xxx | yyy | zzz |	Pour faire un tableau simple il suffit de séparer les colonnes par la barre verticale. Cette méthode autorise uniquement des cellules d'une seule ligne. Pour faire des tableaux plus complexes, utiliser le langage HTML.
Insérer une image	• <IMG1|left> • <CENTER> <IMG1|center> </CENTER> • <IMG1|right>	L'image doit avoir été précédemment téléchargée sur le site par la fonction « télécharger une nouvelle image » dans la colonne de gauche affichée quand on rédige un article. Le système indique alors sous quel nom l'image est disponible, et c'est ce nom qui doit être utilisé dans les commandes ci-contre. Trois commandes sont disponibles, selon qu'on veut voir l'image à gauche, au centre ou à droite.

Pour plus d'informations, dans l'espace d'administration de SPIP, cliquer sur [AIDE], puis «les articles», et enfin « les raccourcis typographiques ».

Annexe 6

Présentation d'une fiche d'information

INIST-CNRS	ARtist	DATE DE CREATION : 14/06/2005
Projet E-BioSci	Fiche d'information	Auteur : Djamila Safa

E-BioSci

Plate-forme européenne pour l'accès à l'information en texte intégral en Sciences de la vie

Le projet E-BioSci

E-BioSci est un nouveau projet pour l'information scientifique apparu sur internet à l'initiative de l'EMBO (European Molecular Organization) soutenu par la commission européenne, pour faire face au problème de la croissance exponentielle des volumes d'informations biologiques. E-BioSci [1] offre aux chercheurs l'accès à un ensemble de données (moléculaires, littérature scientifique, images) dans le domaine des sciences de la vie.

Objectif du projet E-BioSci

E-BioSci (European platform for access and retrieval of full text and factual information in life sciences) est un projet innovant qui offre dans le domaine des sciences de la vie une gamme de services associés à différents types d'informations :

- Recherche croisée en français, en anglais et en allemand
- recherche de concept dans le texte intégral (outil de découverte)
- utilisation des empreintes conceptuelles pour lier du texte de façon sémantique à différents types de données (données génomiques et d'images)
- liens avec des documents validés qui répondent à des critères de contrôle éditorial
- libre accès pour le monde universitaire et de la recherche dans le respect des restrictions imposées par les fournisseurs de données

La plate-forme E-BioSci

Le prototype utilise un système d'« empreintes » de documents, mis au point par Collexis B.V. en s'appuyant sur des ressources comme le thésaurus MeSH, des bases de données telles que Medline, PASCAL, Ingenta, Nature publications, BioMed Central, PubMed Central. Ces dernières permettent l'accès libre en ligne aux revues en texte intégral, aux citations et aux résumés.

L'empreinte conceptuelle

Ce concept permet l'identification des symboles de gène et de protéine dans un texte intégral avec une indication des services de consultation de base de données disponibles.

Les partenaires européens

La dimension européenne d'E-Biosci a permis de rassembler autour de l'EMBO, initiatrice et coordinatrice du projet, sept partenaires européens :

- **CINDOC-CSIC** (Centro de Información y Documentación Científica), http://www.cindoc.csic.es
- **CINES** (Centre Informatique national de l'Enseignement Supérieur), http://www.cines.fr
- **DIMDI** (Deutsches Institut für Medizinische Dokumentation und Information), http://www.dimdi.de
- **EDINA** (Edinburgh University Data Library), http://datalib.ed.ac.uk/
- **EMBL-EBI** (European Bio-informatics Institute), http://www.ebi.ac.uk/
- **Ingenta** UK Ltd, http://www.ingenta.com
- **INIST** (Institut d'information scientifique et technique) www.inist.fr
- Et un sous-traitant **Collexis B.V.** (http://www.collexis.com/)

Note 1

Pour obtenir de plus amples renseignements sur le projet E-BioSci, visiter les sites suivants :

- http://www.e-biosci.org/
- http://artist.inist.fr/article.php3?id_article=118

Annexe 7

Code CSS qui a permis un affichage par défaut des fiches dans le site ARTIST

```
/* Mise en page des Fiches du site Artist*/
#fiche{
border:solid 0.2em #3C7CBD;
padding :1em;
}
/* parametres communs*/
#titrefiche, #soustitrefiche {
text-align:center;
width: 70%;
margin-bottom: 5%;
border : solid 1px #3C7CBD;
}
td {
	font-size: 90%;
	font-weight: bold;
	padding: 0.2em;
	border : solid 1px #3C7CBD;
}
span {
	font-size: 90%;
	font-weight: bold;
	padding: 0.2em;
}
h3.spip {
	background-color: #8FC7FF;
	padding:0.4em;
	margin-top:1em;
	display: table;
	text-align: center;
}
/* style du titre*/
#titrefiche{
	margin-bottom: 5%;
	width: 90%;
}
#titrefiche .cote{
	background: #8FC7FF;
	width: 25%;
}
#titrefiche .centre{
	background: #DCDCDC;
}
/* style du sous-titre*/
#soustitrefiche{
width: 70%;
}
#soustitrefiche .cote{
	background:transparent;
	width: 20%;
}
#soustitrefiche .centre{ background: #8FC7FF;}
```

Annexe 8

Comment rédiger une requête sur PASCAL?

Champs :

```
ab  : Résumés anglais ou français (fa + ea)
an  : Numéro Pascal
au  : Auteurs
cc  : Codes de classement
ch  : Nom du champ
dt1 : Type de document primaire
dt2 : Type de document secondaire
dtm : Mode
ea  : Résumés anglais
ed  : Descripteurs anglais
eg  : Termes génériques anglais
et  : Titres anglais
fa  : Résumés français
fd  : Descripteurs français
fg  : Termes génériques français
ft  : Titres français
kw  : Mots-clés anglais (ed + eg)
la  : Langue du texte
mc  : Mots-clés français (fd + fg)
nb  : Niveau bibliographique
or  : Origine de la notice
pas : Numéro de bande PAS (depuis 1997)
sn  : ISSN
ti  : Titres anglais ou français (et + ft)
```

Sans indication de champ, la requête se fait sur l'index de base qui réunit les champs "ab", "kw", "mc" et "ti".

Conseils d'utilisation :

1 - Ecrivez en caractères minuscules, sans accent, sauf exception signalé plus bas,

2 - Les opérateurs booléens sont "et", "ou" et "sauf",

3 - Utilisez le signe "=" pour indiquer le champ : mc = souris

4 - Parenthésez si nécessaire : mc = (souris ou rat)

5 - Utilisez les guillemets pour les multitermes (champs "au","kw", "mc" et "sn"): mc = "demence alzheimer", sn = "0302-8933"

Par extension, un uniterme entre guillemets représente un descripteur à part entière et non pas un élément d'un descripteur. On peut avoir : mc = (homme sauf "homme")

6 - Les lettres grecques sont verbalisées : mc = "alpha amylase"

7 - La troncature "*" indique aucun ou plusieurs caractères, alors que le masque "?" indique un caractère quelconque,

8 - Il est possible d'utiliser les troncatures en début de mot, mais il faut savoir que cela augmente fortement le temps de réponse ; donc, utilisez cette fonctionnalité avec parcimonie,

9 - Le champ "et" doit être précédé d'une barre inverse "\" pour éviter la confusion avec l'opérateur booléen "et". De même, tout élément d'une requête pouvant être confondu avec un opérateur booléen doit être protégé par une barre inverse.

Cas particulier :

1 - Champ "an" : la présence des guillemets est indispensable. Les requêtes sont de la forme : an = "92-0315724"

2 - Champ "dt1", "dt2", "dtm", "la", "nb" et "or" : ces champs sont indexés avec les codes du format INIST. Les requêtes sont de la forme :

dt1 = c et nb = a
la = (esp ou ita)
or = 02*

3 - Champ "pas" : toutes les entrées de cet index ont 2 caractères et les requêtes sont de la forme : pas = (01 ou 02 ou 03)

4 - Champ "ch" : toutes les entrées de cet index ont 12 caractères et les requêtes sont de la forme : ch = A09d025ENGs1

Ceci signifie qu'une requête comme "ch = A09" ou
"ch = A09d???ENG" n'aura jamais aucun résultat.
Donc, n'oubliez pas la troncature : ch = A09d???ENG*

Certains champs ont de nombreuses entrées dans cet index : "C03" en a environ un millier. La requête : ch = C03* prendra environ 1/4 d'heure, alors que la requête : ch = C03d01* ne prendra qu'une minute. A noter que pour Pascal 1992, la meilleure requête, c'est-à-dire la plus complète et la plus rapide, est : ch = C03d0??FRE*
De même, la meilleure requête pour avoir l'ensemble des notices est : ch = A47*

Ma requête sur PASCAL :

(cc= 002* ou mc= ("biologie" ou "biologie cellulaire" ou "biologie moleculaire" ou "sciences vie") sauf (cc= (002b18h05b* ou 002a26*) ou mc= "sante publique" ou "sante mentale" ou "psychotherapie" ou "psychologie" ou "soin" ou "socio*" ou "psychiatrie" ou "attitude" ou "maladie mentale" ou "arrieration mentale" ou "sante" ou " medecine clinique" ou "soin" ou "radiologie medicale" ou "reanimation" ou "soin intensif" ou "education sante" ou "cognition" ou "chirurgie" ou "autisme" ou "comportement" ou "psychanalyste" ou "aspect juridique" ou "physique" ou "demographie" ou "handicap" ou "relations sexuelles" ou "telemedecine" ou "radiologie" ou "imagerie medicale" ou "sciences humaines" ou " dossier medical" ou "Histoire" ou "Secret medical" ou "medecine travail" ou "electrodiagnostic" ou "peur" ou "reve" ou "deuil" ou "foret" ou "environnement" ou "angoisse anxiete" ou "prevention" ou "social" ou "Relation soignant soigne" ou "relation medecin malade")) et (mc = ("information scientifique technique" ou "langage xml" ou " nouvelle technologie information communication" ou "archives ouvertes" ou "communication scientifique" ou "infometrie" ou "litterature scientifique" ou "base donnee" ou "ressource internet" ou "internet" ou "source information" ou "information numerique" ou "metadonnee" ou "systeme information" ou "information reseau" ou "Revue bibliographique" ou "edition electronique" ou "revues

scientifiques" ou "revues numeriques" ou "transfert technologie" ou "communication information" ou "information electronique" ou "information biomedicale" ou "technologie information" ou "acces libre" ou "extraction information" ou "diffusion information" ou "recherche information" ou "analyse information" ou "bibliometrie" ou "information biologique" ou "archive electronique" ou "isi") sauf mc= ("bioinformatique" ou "informatique" ou "geographie" ou "sol" ou "telecommunication" ou "teleinformatique" ou "cout" ou "telematique"))

Ma requête sur Pubmed :

(technical [All Fields] AND scientific[All Fields] OR "scientific journal"[Title] AND information[All Fields] OR "information"[Title]) OR "Review Literature"[MeSH] OR "Utilization Review"[MeSH] OR "numerical information"[All Fields] OR "journal"[Title] OR "journal information"[All Fields] OR "electronic publication"[All Fields] OR "academic information"[All Fields] OR "bibliographic database"[Title] OR "bibliographic databases"[Title] OR "Information Displays"[All Fields] OR "scientific information"[Title] OR "information tools"[Title] OR "scientific literature"[Title] OR "Unified Medical Language System"[All Fields] OR "MEDLARS"[All Fields] OR "information dissemination"[Title] OR "Information Theories"[All Fields] OR "information system"[All Fields] OR "information system"[Text word] OR "academies and institutes"[MeSH Terms] OR "exploring information"[Title] OR ("information retrieval"[All Fields] AND storage[All Fields]) OR "Access to Information"[Title] OR "information technologies"[Title] OR "communication information"[Title] OR "extraction information"[Title] OR "information technologies"[All Fields] OR "internet"[Title] OR "Integrated Advanced Information Management Systems"[All Fields] OR "bibliometric analysis"[All Fields] OR "biomedical information"[All Fields] OR "Metadata"[All Fields] OR "publishing"[Title] OR "reviews"[Title] OR "Technology Transfer"[MeSH] OR "Technology Transfer"[Title] NOT ("informatics"[All Fields] OR "bioinformatics"[All Fields] OR "public health"[All Fields] OR "dental"[All Fields] OR "surgical"[All Fields] OR "medicine"[All Fields] OR "telemedicine"[All Fields] OR "adolescent sexual"[All Fields] OR "sexual education"[All Fields] OR "psychology"[All Fields] OR "mental health"[All Fields] OR "Psychological barriers"[Title] OR "anthropology"[All Fields] OR "Social Sciences"[MeSH] OR "Health Care Category"[Mesh] OR "dentistry"[All Fields] OR "health care"[Title] OR "Health Occupations"[Mesh] OR "Emergency Medical Services"[Mesh] OR "diabetes"[All Fields] OR "drug"[All Fields] OR "ubiquitin"[Title] OR "deases"[All Fields] OR "radiology"[All Fields] OR "Physical Therapy Techniques"[All Fields] OR "Hospital admissions"[All Fields]) AND "1995/07/29 05.46"[EDAT] : "2005/07/26 05.46"[EDAT].

Annexe 9

1) Format serveur d'un corpus :

Une notice en est constituée d'un ensemble structuré de champs et chacun d'entre eux contenant une information typée. L'exemple suivant montre un exemple de notice utilisée dans mon projet.

NO : PASCAL 05-0242666 INIST
ET : Co-occurrence based meta-analysis of scientific texts : retrieving biological relationships between genes
AU : JELIER (R.); JENSTER (G.); DORSSERS (L. C. J.); VAN DER EIJK (C. C.); VAN MULLIGEN (E. M.); MONS (B.); KORS (J. A.)
AF : Department of Medical Informatics, Erasmus MC-University Medical Center/Rotterdam/Pays-Bas (1 aut., 4 aut., 5 aut., 6 aut., 7 aut.); Department of Urology, Erasmus MC-University Medical Center/Rotterdam/Pays-Bas (2 aut.); Department of Pathology, Erasmus MC-University Medical Center/Rotterdam/Pays-Bas (3 aut.)
DT : Publication en série; Niveau analytique
SO : Bioinformatics : (Oxford. Print); ISSN 1367-4803; Royaume-Uni; Da. 2005; Vol. 21; No. 9; Pp. 2049-2058; Bibl. 45 ref.
LA : Anglais
EA : Motivation: The advent of high-throughput experiments in molecular biology creates a need for methods to efficiently extract and use information for large numbers of genes. Recently, the associative concept space (ACS) has been developed for the representation of information extracted from biomedical literature. The ACS is a Euclidean space in which thesaurus concepts are positioned and the distances between concepts indicates their relatedness. The ACS uses co-occurrence of concepts as a source of information. In this paper we evaluate how well the system can retrieve functionally related genes and we compare its performance with a simple gene co-occurrence method. Results: To assess the performance of the ACS we composed a test set of five groups of functionally related genes. With the ACS good scores were obtained for four of the five groups. When compared to the gene co-occurrence method, the ACS is capable of revealing more functional biological relations and can achieve results with less literature available per gene. Hierarchical clustering was performed on the ACS output, as a potential aid to users, and was found to provide useful clusters. Our results suggest that the algorithm can be of value for researchers studying large numbers of genes.
CC : 002A01B
FD : Métaanalyse; Gène; Criblage haut débit; Biologie moléculaire; Méthode; Utilisation; Nombre; Information biomédicale
ED : Metaanalysis; Gene; High throughput screening; Molecular biology; Method; Use; Number; Biomedical information
SD : Meta-análisis; Gen; Cribado alta productividad; Biología molecular; Método; Uso; Número; Información biomédical
LO : INIST-21331.354000125440040380

Légende : copie d'écran d'une notice PASCAL

2) Mode opératoire de déchargement des corpus (PASCAL et Pubmed) sur Leximine

Le découpage des corpus en notices est effectué avec une invite de commandes DOS des bases de données PASCAL et Pubmed. Pour cela, on procède comme suit :

- changer de lecteur : "g:"
- aller au bon répertoire où se trouve le programme : « cd Mine\sbin\mine »
- la commande pour lancer le programme :
  - "lqmine -c o:\djamila\Corpus.txt NO: o:\djamila\Notices" (pour PASCAL)

"lqmine -c o:\djamila\Medline\Pubmed-result.txt PMID: o:\djamila\Medline\Notices" (pour Pubmed).

Annexe 10

Requête établie sur Keywatch est la suivante :

(Scientific* technic* information*) or Database* or Bibliometri* or (Technology transfer) or (biological information*) or (Biomedical information *) or UMLS or (Information technolog*) or (Information source*) or (scientific journal*) or (Scientific literature*) or (Electronic information*) or (published literature*) or (scientific* Review*) or (published Review*) or (bibliographic review*) or (open access) or (Information extraction) or (web site*) or metadata* or (Scientific* communication*) or (Information retrieval) or Documentation

Liste des figures et tableaux

Liste des figures

FIGURE 1 : Répartition par domaines du fonds documentaire des périodiques (26000 titres) a l'INIST 8
FIGURE 2 : Présentation de générale du site ARTIST 13
FIGURE 3 : Représentation schématique du site ARTIST et de son contenu 15
FIGURE 4 : Présentation de l'interface du site ARTIST (partie publique) 18
FIGURE 5 : L'espace prive dans ARTIST 18
FIGURE 6 : Affichage d'une fiche d'information dans le site ARTIST 24
FIGURE 7 : Représentation d'une arborescence de termes Mesh pour le domaine des sciences de l'information 27
FIGURE 8 : Interface d'une représentation cartographique sur Leximine 29
FIGURE 9 : Traitement des notices avant le découpage du corpus (exemple d'une notice Pubmed) 30
FIGURE 10 : Module Administration dans Leximine 31
FIGURE 11 : Interface flash de l'utilisateur dans Keywatch 32
FIGURE 12 : Présentation du menu « source » dans Keywatch 32
FIGURE 13 : Distribution des mots-clés dans le corpus édutié (BDD PASCAL) 33
FIGURE 14 : Distribution par pays de publication dans le corpus étudié 34
FIGURE 15 : Carte montrant les associations entre les termes sur Leximine 35
FIGURE 16 : Interface complète d'une cartographie Leximine 36
FIGURE 17 : Carte montrant la sous-thématique « immunologie » et tous les termes qui lui sont associés 37
FIGURE 18 : Carte montrant un organisme travaillant sur le sujet traité 37
FIGURE 19 : Cartographie montrant les associations autour du terme « Metadata » issu du corpus étudié (BDD PASCAL) sur Leximine 38
FIGURE 20 : Distribution par pays de publication (PASCAL sur Leximine) dans le corpus étudié 40

FIGURE 21 : Distribution par pays de publication (BDD PASCAL sur Leximine) 41
FIGURE 22 : Distribution des termes dans le corpus étudié sur Leximine (BDD Pubmed) 42
FIGURE 23 : Carte montrant les associations autour du terme « Internet » 43
FIGURE 24 : Distribution par pays de publication dans le corpus étudié (BDD Pubmed sur Leximine)..... 444
FIGURE 25 : Distribution par pays de publication (cartographie Leximine de la BDD Pubmed) 45
FIGURE 26 : A gauche affichage des URL de la base créée sous mon compte. A droite les nouveautés identifiées des pages web..... 46
FIGURE 27 : Carte de visualisation des associations autour du terme « UMLS » sur Leximine de la base créée sur Keywatch 47

Liste des Tableaux

TABLEAU 1 : Présentation de quelques résultats d'analyse autour du terme « metadata » issu du corpus PASCAL sur Leximine..... 39
TABLEAU 2 : Présentation de quelques résultats d'analyse autour du pays « Afrique » issu du corpus PASCAL sur Leximine 41
TABLEAU 3 : Présentation de quelques résultats d'analyse autour du terme « Internet » issu du corpus Pubmed sur Leximine 43

Résumé

ARTIST est une initiative sur l'appropriation par la recherche des technologies de l'IST et de la communication scientifique qui commence à trouver ces marques. C'est aussi un projet pilote à trois volets : recherche (animation de groupe de travail sur cette thématique avec la création d'une revue scientifique électronique), veille scientifique et technologique (un observatoire avec une approche thématique) et publication en ligne (un site web géré par SPIP). Il s'intéresse à un large panorama d'activités : édition électronique, constitution de bases de données, activités terminologiques et opérations innovantes dans lesquelles un mécanisme d'appropriation par la communauté scientifique est clairement identifié. Au cours de mon stage, j'ai participé à la structuration du site par l'adaptation du script dans SPIP, mais également à la gestion de son contenu en publiant deux articles dans l'espace jeunes chercheurs et en créant et diffusant des fiches d'information de projets dans la rubrique « Sciences de la vie et Médecine ». Une étude de veille sur le thème de l'appropriation des technologies de l'IST dans le domaine biomédical a été réalisée pour apporter d'autres informations complémentaires dans le but d'alimenter le site.

Avec ARTIST, l'INIST entre dans une nouvelle expérience de travail coopératif avec le monde de la recherche.

Mots-clés : Projet ARTIST, information scientifique technique, information biomédicale, observatoire de technologies, transfert de technologies, bibliométrie, édition électronique, NTIC, édition coopérative, libre accès, sciences de la vie.