

E.N.S.S.I.B.
ECOLE NATIONALE SUPERIEURE
DES SCIENCES DE L'INFORMATION
ET DES BIBLIOTHEQUES

UNIVERSITE
CLAUDE BERNARD
LYON I

DESS en INFORMATIQUE DOCUMENTAIRE

Rapport de stage

**Histoire d'une base de données
qui voulait voyager sur Internet**

~

**NosoBase,
base de données bibliographiques du C.CLIN Sud-Est**

Jérôme GUTMAN

Sous la direction de

Louis AYZAC

*Praticien Hospitalier Universitaire
HEIM - Université Lyon 1*

Nathalie LEROY

*Documentaliste
C.CLIN Sud-Est*

Jean-Pierre LARDY

*Maître de conférences
URFIST - Université Lyon 1*

Année 1995-1996

Histoire d'une base de données qui voulait voyager sur Internet

~

NosoBase, base de données bibliographiques du C.CLIN Sud-Est

Jérôme GUTMAN

RESUME

Ce document retrace quatre mois de stage effectués au C.CLIN Sud-Est (Centre de Coordination de la Lutte contre les Infections Nosocomiales). Il s'agit de la mise à disposition d'une base de données bibliographiques, NosoBase, sur Internet, en tant que base WAIS : reformatage des données, construction de la base (FreeWAIS-sf), du formulaire d'interrogation (SFgate) et d'un serveur WWW

DESCRIPTEURS¹

BASE DONNEE BIBLIOGRAPHIQUE, INFECTION NOSOCOMIALE, INTERNET, SERVEUR INFORMATIQUE, WAIS (WIDE AREA INFORMATION SERVER)

ABSTRACT

This document reports on a four-month probationary period at C.CLIN Sud-Est (coordination center for hospital infection prevention and control in the south-east of France). It is about the setting up of a bibliographic database, NosoBase, that has been put on the Internet as a WAIS database : data reformatting, database (FreeWAIS-sf), query form (SFgate) and WWW server construction are reported.

KEYWORDS

BIBLIOGRAPHIC DATABASE, COMPUTER SERVER,

¹ D'après le CD-ROM 1995 de la base Pascal (INIST). La liste des descripteurs est aussi disponible dans : INIST. - Pascal Lexique [Sciences exactes et technologies. Sciences de la vie. Lexique français-anglais-espagnol]. - Vandoeuvre-lès-Nancy, ed. INIST, 1991. - ISBN 2904975638. - Bibliothèque de l'ENSSIB : côte ENSSIB U025.49 CEN

INTERNET, NOSOCOMIAL INFECTION, WAIS

Table des matières

TABLEAUX, GRAPHIQUES, FIGURES, IMAGE.....	7
INTRODUCTION.....	8
I PRÉSENTATION DU STAGE.....	9
1. LES CENTRES DE COORDINATION DE LA LUTTE CONTRE LES INFECTIONS NOSOCOMIALES (C.CLIN)9	
1. 1. Statut.....	9
1. 2. Missions.....	10
1. 3. Les différents sites.....	11
2. LE C.CLIN SUD-EST.....	11
2. 1. Activités.....	12
2. 2. Structure et personnel.....	14
2. 3. Moyens informatiques.....	15
2. 3. 1. L'unité d'Hygiène, Epidémiologie et Information Médicale (HEIM).....	15
2. 3. 2. Réseaux.....	15
2. 3. 3. Matériel.....	16
2. 3. 4. Logiciels.....	16
3. LE CENTRE DE DOCUMENTATION DU C.CLIN SUD-EST : NOSOBASE	17
3. 1. Personnel.....	17
3. 2. Public.....	18
3. 3. Fonds documentaire.....	18
3. 4. Activités.....	21
3. 4. 1. Sélection des articles et indexation.....	21
3. 4. 2. Saisie des notices.....	21
3. 4. 3. Accueil du public.....	21
3. 4. 4. Interrogation de NosoBase.....	21
3. 4. 4. 1. Accès indirect illimité.....	21
3. 4. 4. 2. Accès direct limité.....	22
3. 4. 4. 3. Statistiques des demandes au C.CLIN Sud-Est pour l'année 1995.....	22
3. 4. 5. Fourniture de documents.....	23
3. 4. 6. Produits documentaires.....	23
3. 4. 7. La base de données NosoBase.....	24
3. 4. 7. 1. Le logiciel documentaire.....	24
3. 4. 7. 2. Les notices.....	25
3. 4. 8. Moyens informatiques.....	25
3. 4. 9. Coopération inter-C.CLIN.....	26
4. OBJECTIF DU STAGE.....	26
II MISE À DISPOSITION DE NOSOBASE SUR INTERNET :.....	27
1. CAHIER DES CHARGES	27
1. 1. Faisabilité du projet.....	27
1. 1. 1. Les données du problème : réseau privé vs Internet.....	27

1. 1. 2. Coûts.....	28
1. 1. 3. Motivations et Conduite du projet.....	29
1. 1. 4. Conditions matérielles.....	29
1. 1. 4. 1. Connection.....	29
1. 1. 4. 2. Logiciel.....	30
1. 2. Objectifs.....	30
1. 3. Répartition des tâches entre les C.CLIN.....	31
1. 4. Contenu.....	31
1. 5. Utilisateurs.....	32
1. 6. Calendrier et étapes du travail.....	32
2. CRÉATION D'UN PROGRAMME DE SAISIE DES NOTICES EN LOCAL.....	34
2. 1. Circuit des données.....	34
2. 2. Choix du langage WordBasic.....	34
2. 2. 1. Avantages.....	34
2. 2. 2. Inconvénients.....	35
2. 3. Format à obtenir.....	36
1. 1. Le programme en macros-commandes Word.....	39
1. 1. 1. NOSOInstall : macro à exécuter uniquement à la première installation.....	40
1. 1. 2. NOSOMenu : menu général.....	41
1. 1. 3. NOSOSaisie : saisie des notices.....	41
1. 1. 4. NOSOCompte.....	42
1. 1. 5. NOSONouvelleplagenotices.....	42
1. 1. 6. NOSOPréparenvoi.....	42
1. 1. 7. FTP.....	43
1. 2. Matériel fourni.....	43
1. 3. Test.....	43
2. NOSOBASE SUR INTERNET.....	44
2. 1. Choix du moteur de recherche et du serveur.....	44
2. 1. 1. Le moteur de recherche : WAIS.....	44
2. 1. 2. Le serveur du Centre d'Information Scientifique et Médicale (CISM) de l'Université Lyon	
2. 1. 2.....	45
2. 2. WAIS.....	46
2. 2. 1. Le passé.....	46
2. 2. 1. 1. Présentation.....	47
2. 2. 1. 2. La norme Z39.50.....	47
2. 2. 1. 3. Les logiciels Wais.....	47
2. 2. 1. 4. Les bases WAIS.....	48
2. 2. 1. 5. Interrogation des bases.....	48
2. 2. 2. Le présent.....	50
2. 2. 2. 1. WAIS existe toujours.....	50
2. 3. FreeWAIS-sf.....	51
2. 3. 1. Installation.....	51
2. 3. 2. Le logiciel d'indexation.....	52

2. 3. 3. Syntaxe d'interrogation.....	53
2. 3. 3. 1. Champs :.....	53
2. 3. 3. 2. Opérateurs.....	53
2. 4. WWW-WAIS : SFgate.....	54
2. 4. 1. Installation.....	55
2. 4. 2. La création de formulaires.....	55
2. 4. 3. Syntaxe d'interrogation.....	56
2. 4. 4. Visualisation des résultats.....	56
2. 5. Mise en place des données et du moteur sur le serveur.....	56
2. 5. 1. Création de la base WAIS.....	56
2. 5. 1. 1. Description des notices.....	57
2. 5. 1. 2. Le reformatage des données vers WAIS.....	58
2. 5. 1. 2. 1. Macros-commandes Word.....	58
2. 5. 1. 2. 2. Clipper, langage de Microsoft Foxpro.....	59
2. 5. 1. 3. Indexation des fichiers.....	60
2. 5. 1. 4. Le fichier .src.....	61
2. 5. 2. L'interface d'interrogation SFgate.....	61
2. 5. 2. 1. Le formulaire HTML/SFgate.....	61
2. 5. 2. 1. 1. Interrogation avancée.....	62
2. 5. 2. 1. 2. Interrogation simple.....	65
2. 5. 2. 2. Exemple d'interrogation avancée.....	66
2. 5. 2. 3. Un produit perfectible.....	66
2. 5. 3. Le serveur WWW NosoBase.....	68
2. 5. 3. 1. Développement HTML.....	68
2. 5. 3. 1. 1. Editeurs utilisés.....	68
2. 5. 3. 1. 2. Particularité HTML : les frames.....	69
2. 5. 3. 1. 2. 1. Syntaxe.....	70
2. 5. 3. 1. 2. 2. Exemple.....	71
2. 5. 3. 2. "Philosophie" du serveur.....	73
2. 5. 3. 3. Contenu.....	73
2. 5. 3. 4. Lancement du serveur.....	75
2. 5. 3. 4. 1. Ouverture au public.....	75
2. 5. 3. 4. 2. Une première phase de test avant la promotion.....	76
2. 5. 3. 5. Une charge de travail supplémentaire.....	77
2. 6. Saisir les données par Internet : l'interface CGI.....	77
2. 6. 1. Principes CGI.....	77
2. 6. 2. le langage Perl.....	78
2. 6. 3. Appel du script : formulaire et droits d'accès.....	79
3. PROTECTION DE "NOSOBASE" À L'INPI.....	80
3. 1. Protéger ses noms : où ?.....	80
3. 2. Comment ?.....	80
3. 2. 1. Recherche d'antériorité.....	81
3. 2. 2. Dépôt.....	83

3. 3. Quand ?.....	83
3. 4. Pourquoi ?.....	84
3. 5. En résumé.....	84
3. 6. Avancée des travaux.....	86
4. PERSPECTIVES POUR NOSOBASE.....	86
CONCLUSION.....	88
BIBLIOGRAPHIE.....	89
BOOKMARK : ADRESSES INTERNET.....	91
Liste des abréviations utilisées.....	97
INDEX.....	98
ANNEXES.....	101

Tableaux, graphiques, figures, image

Tableau 1 : définition des infections nosocomiales.....	9
Tableau 2 : répartition géographique des cinq C.CLIN.....	11
Tableau 3 : équipe du C.CLIN sud-est.....	14
Graphique 1 : utilisateurs de NosoBase par catégories professionnelles.....	18
Graphique 2 : types de documents dans NosoBase.....	20
Graphique 3 : dates de publication des documents.....	20
Graphique 4 : langue des documents.....	20
Graphique 5 : nombres de demandes 1994/1995 au centre de documentation.....	23
Figure 1 : le réseau privé des HCL.....	27
Tableau 4 : la solution proposée par la DSII.....	28
Tableau 5 : répartition des tâches entre les C.CLIN.....	31
Tableau 6 : circuit des données Basis plus / Internet.....	34
Tableau 7 : champs Basis plus.....	38
Figure 2 : programme de saisie en local.....	40
Image 1 : boîte de dialogue du masque de saisie.....	40
Figure 3 : fonctionnement de SFgate.....	54
Tableau 8 : champs Wais de NosoBase.....	58
Graphique 6 : les frames dans departf.htm.....	71
Tableau 9 : tarif général de dépôt à l'INPI.....	83
Tableau 10 : tarifs pour une recherche d'antériorité de NosoBase à l'INPI.....	85
Tableau 11 : tarifs pour un dépôt de NosoBase.....	85
Tableau 12 : coût total pour la protection à l'INPI de NosoBase.....	86

Introduction

Il était une fois une base de données qui voulait voyager sur Internet...

NosoBase, base de données bibliographiques sur les infections nosocomiales (*se dit d'une infection contractée lors d'un séjour en milieu hospitalier*), a été créée par le Centre de Coordination de la Lutte contre les Infections Nosocomiales du Sud-Est (C.CLIN Sud-Est), aux débuts des années 90.

Jusqu'alors isolée, dans le sens où personne ne pouvait la consulter à distance, les cinq C.CLIN de France décidèrent en 1996 de la rendre disponible aux six coins de l'hexagone, voire plus loin, dans les steppes de Sibérie, au beau milieu de l'Amazonie...

Une aventure de quatre mois (Juin - Septembre 1996) m'a permis de préparer ce voyage, à travers plusieurs épisodes :

- la création d'un programme de saisie des notices pour tous les C.CLIN
- le reformatage des données du logiciel documentaire Basis plus
- la construction de *NosoBase* en tant que base WAIS (standard documentaire sur Internet), sous FreeWAIS-sf
- son interrogation par SFGate (passerelle pour interroger la base WAIS) à l'aide de formulaires HTML
- sa présentation sur les pages d'un serveur Internet WWW
- les débuts d'une saisie des notices à distance (script CGI)
- enfin, la protection de *NosoBase*, en tant que marque déposée

Après avoir présenté le C.CLIN Sud-Est et sa base documentaire, nous décrirons donc ces différentes étapes.

Alors, en avant pour le voyage...

Note : vous pourrez trouver dans les parties Bibliographie (page 89) et Bookmark (page 91) les références d'ouvrages, de documents ou de serveurs Internet qui sont mentionnés par la suite, ainsi qu'une liste des abréviations utilisées (page 97).

I Présentation du stage

1. Les Centres de Coordination de la Lutte contre les Infections Nosocomiales (C.CLIN)

<i>NOSOCOMIAL, E, AUX adj. (gr. nosos, maladie, et komeîn, soigner)</i>
Se dit d'une infection contractée lors d'un séjour en milieu hospitalier
Le Petit Larousse Illustré, 1992

Tableau 1 : définition des infections nosocomiales

1. 1. Statut

Les infections hospitalières constituent un problème majeur de Santé Publique. L'obligation d'organiser la lutte contre les infections nosocomiales existe depuis le décret N°88-657 du 6 Mai 1988 relatif à l'organisation de la surveillance et de la prévention des infections nosocomiales dans les établissements d'hospitalisation publics et privés participant au service public hospitalier. Ce texte définit les missions des Comités de Lutte contre les Infections Nosocomiales (CLIN), présents dans les organismes de soins. Il recommande "d'organiser et de coordonner une surveillance continue des infections dans les établissements" et de "promouvoir des actions de formation des personnels". Ceci consiste en la mise en oeuvre d'une surveillance épidémiologique de l'infection, de l'évaluation de la qualité des soins dispensés dans l'établissement et en la mise en place d'un programme de prévention.

L'arrêté du 3 Août 1992 crée le Comité Technique National des Infections Nosocomiales (CTIN) et les 5 Centres de Coordination pour la Lutte contre les Infections Nosocomiales (C.CLIN). Le CTIN a comme objectif prioritaire de définir le "programme minimum de surveillance des infections nosocomiales". Il est chargé de formuler les recommandations techniques et il synthétise dans un rapport national

annuel l'activité des cinq C.CLIN. Ces centres disposent de moyens de fonctionnement et de personnel pour améliorer l'organisation de la lutte contre les infections nosocomiales dans l'inter-région et mener à bien des actions adaptées aux priorités nationales et locales.

1. 2. Missions

Les C.CLIN font donc partie du service public. Leurs missions sont définies par l'article 7 de l'arrêté du 3 Août 1992 :

1. "La coordination des actions de lutte contre les infections nosocomiales conduites par les établissements ou organismes publics ou privés de prévention ou de soins"
2. "L'organisation du recueil épidémiologique standardisé de données d'incidence et de prévalence des infections nosocomiales à partir d'un réseau local d'établissements"
3. "La mise au point de protocoles d'investigations épidémiologiques"
4. "La réalisation d'études épidémiologiques multicentriques sur les risques infectieux et leur prévention"
5. "L'élaboration et la mise à jour d'un guide de l'hygiène et des pratiques de soins"
6. "L'assistance technique aux établissements lors de la survenue de cas groupés et pour leur programme de formation en matière de lutte contre les infections nosocomiales"
7. "La constitution et la mise à jour d'un annuaire des prestataires de services dans le domaine de l'hygiène hospitalière et des sociétés sous-traitantes pour les établissements"
8. "La tenue d'un fichier documentaire et bibliographique"
9. "L'animation d'un réseau de responsables de CLIN"
10. "L'assistance technique et le conseil aux établissements qui en formulent la demande".

1. 3. Les différents sites

Les cinq C.CLIN sont répartis de la façon suivante :

Situation géographique	Régions concernées
C.CLIN Est : Strasbourg	Alsace, Ardennes, Bourgogne, Champagne, Franche-Comté, Lorraine
C.CLIN Ouest : Rennes	Basse Normandie, Bretagne, Centre, Pays de la Loire
C.CLIN Paris-Nord : Paris	Haute Normandie, Ile de France, Picardie, Nord-Pas de Calais
C.CLIN Sud-Est : Pierre-Bénite	Auvergne, Corse, Ile de la Réunion, Languedoc Roussillon, Provence Alpes Côte d'Azur, Rhône Alpes
C.CLIN Sud-Ouest : Bordeaux	Antilles, Aquitaine, Guyane, Limousin, Midi Pyrénées, Poitou Charentes

Tableau 2 : répartition géographique des cinq C.CLIN

2. Le C.CLIN Sud-Est

Chargé de tout le quart Sud-Est de la France, y compris l'Ile de la Réunion, le C.CLIN Sud-Est fait partie du grand ensemble des Hospices Civils de Lyon (HCL). Il se situe au Centre Hospitalier Lyon Sud à Pierre-Bénite (CHLS, anciennement Jules Courmont et St Eugénie), au Pavillon 1 M, au milieu d'espaces boisés, d'immeubles de soins hospitaliers (le CHLS est particulièrement spécialisé en cancérologie, appareil locomoteur et gériatrie), en dessous de l'Université de Médecine Lyon-Sud.

Le bâtiment qui l'héberge au sous-sol abrite aussi au rez-de-chaussée et au premier étage le service d'Hygiène, Epidémiologie et Information médicale (HEIM). Pas de patients dans le bâtiment, donc, puisque les missions sont d'ordre informatif et préventif.

La coopération et l'entente entre ces différents services sont propices, d'autant qu'ils sont sous la responsabilité d'un même chef de service, partagent le même réseau informatique, des salles de réunion et de repos communes et oeuvrent tous dans un domaine proche.

2. 1. Activités

Les objectifs des C.CLIN sont réalisés par le site Sud-Est au travers des points suivants :

- Information et animation de l'ensemble des CLIN du Sud-Est avec un rôle d'assistance et de conseil auprès des établissements publics ou privés. Envoi trimestriel de documentation aux présidents de CLIN
- Organisation de formations mises à la disposition des hospitaliers, membres du CLIN ou impliqués dans un projet d'hygiène
- Elaboration de guides pratiques et recommandations : *Guide Technique d'Hygiène hospitalière*, participation à la création d'un CD-ROM (*Nosocomix* : infections nosocomiales)
- Diffusion d'un annuaire de ressources Sud-Est
- Réalisation d'études multicentriques sur les risques infectieux et leur prévention
- L'organisation d'un recueil épidémiologique multicentrique des données d'incidence à partir de réseaux d'établissements : réseaux de surveillance des Infections du Site Opératoire (ISO Sud-Est), des Bactériémies Nosocomiales, des Infections Nosocomiales en REAnimation (REA Sud-Est). Ces réseaux fonctionnent en continu et sont en permanence ouverts à de nouveaux participants.
- Participation à l'Enquête Nationale de Prévalence (recensement des infections nosocomiales au plan national)
- Tenue d'une base de données documentaire sur les Infections Nosocomiales, NosoBase, ce qui nous concerne plus particulièrement (voir page 17).

Par ailleurs, certains médecins participent à des congrès et le Professeur J. FABRY est rédacteur en chef de la revue médicale *HygièneS*.

2. 2. Structure et personnel

La structure du C.CLIN Sud-Est s'organise ainsi :

- Assemblée annuelle des CLIN, où sont prises en compte les demandes des hospitaliers dans le domaine des responsabilités du centre (généralement en Octobre ou Novembre)
- Conseil scientifique (Décembre) qui oriente et évalue les activités du centre.

L'équipe du C.CLIN Sud-Est est composée de :

Nom	Statut
J. FABRY	Directeur scientifique
A. SAVEY	Coordonnateur
D. MONNET	Coordonnateur
G. PINZARU	Médecin attaché
M. AUROY	Médecin attaché
C. JACQUEMIER	Médecin attaché
M. COIRON	Cadre Infirmier
N. LEROY	Documentaliste
R. TERNES	Secrétaire

Tableau 3 : équipe du C.CLIN Sud-Est

Le Professeur J. FABRY est donc à la tête du service, assisté des deux médecins coordonnateurs, A. SAVEY et D. MONNET. En dehors de ces trois personnes, aucune hiérarchie ne se fait sentir au travers de l'équipe.

Une réunion à lieu tous les quinze jours pour discuter des projets, soumettre des idées, préparer des événements, etc..., et notamment rencontrer en groupe le professeur FABRY.

Tous les employés possèdent le statut de fonctionnaire dans la majorité.

Enfin, il existe une antenne régionale à Grenoble.

2. 3. Moyens informatiques

2. 3. 1. L'unité d'Hygiène, Epidémiologie et Information Médicale (HEIM)

Comme nous l'avons déjà mentionné, le C.CLIN Sud-Est se trouve dans le même pavillon que le service d'Hygiène, Epidémiologie et Information Médicale, où se trouve le Docteur Louis AYZAC (Praticien Hospitalier Universitaire), informaticien. Son travail nécessite l'emploi important des ordinateurs, et il est donc la principale (et indispensable !) personne ressource en informatique pour tout le bâtiment, y compris le C.CLIN Sud-Est : administrateur du réseau local, il gère l'accès à Internet, conseille pour les commandes de matériel et logiciel, et met à la disposition du personnel hospitalier de Lyon-Sud ses machines, ses logiciels et son savoir-faire.

L'HEIM a été au moins aussi importante dans ce stage (apport informatique) que la partie C.CLIN Sud-Est (apport documentation), c'est pourquoi nous les mentionnerons tous les deux par la suite, sans faire de distinction.

2. 3. 2. Réseaux

L'ensemble du pavillon 1 M est connecté à Internet par faisceau hertzien (2 Mbits/s) vers ROCAD (Reseau Optique du CAmpus de la Doua).

Le bâtiment dispose d'un réseau local Novell : Netware 3.12 50 users, sur machine Compaq (deux disques durs de 500 Mo et 2 Go).

Quelques postes sont équipés pour consulter à distance (TCP/IP) les CD-ROM de la base Medline à la Bibliothèque de Médecine de l'Université Lyon 1.

Enfin, trois machines du C.CLIN Sud-Est sont connectées au réseau des Hospices Civils de Lyon (64 ou 128 Kbit/s) pour l'interrogation de NosoBase uniquement (pas de connexion à Internet ni au réseau Novell : voir page 27).

2. 3. 3. Matériel

Le parc des monopostes est assez hétérogène et plutôt ancien. On peut trouver des Macintosh (Centris 610 et 660 AV, Power PC 7100/66, Performa 450) et des PC

(80486 SX-25 et DX2-66, Pentium 120). Cinq machines environ disposent de lecteurs de CD-ROM.

Le poste le plus performant est une machine HP sous Windows N.T. 3.5.1 et Windows 95 (Pentium Pro 150 Mhz : un monstre de puissance !).

L'HEIM met à la disposition du personnel du CHLS un scanner à plat (couleur 1200 dpi) ainsi qu'une imprimante vidéo Polaroid, qui permet de photographier sur pellicule couleur des pages-écrans (pour projeter sur diapositives des diaporamas réalisés par logiciel de PréAO). Un de leurs postes permet aussi d'interroger les CD-ROM de la base Medline (serveur de CD-ROM à la Bibliothèque de Lyon 1).

Le Professeur FABRY se sert d'un portable, qu'il peut connecter au réseau local et à Internet de l'extérieur.

Tous les postes sont reliés à des imprimantes laser noir et blanc Apple Laserwriter Pro ou Hewlett Packard.

Aucun ordinateur ne m'était personnellement dédié, mais le DX2-66 était le plus souvent libre. La possibilité de stocker les données sur les disques du réseau local permettait de changer de poste facilement.

2. 3. 4. Logiciels

On peut trouver les systèmes d'exploitation suivant : Novell netware 3.12 50 users, MS-DOS 6, Windows 3.1 et 95, Windows N.T. 3.5.1 et MacOS 7.x.

Toutes les machines sont équipées de packs bureautiques Microsoft Office (Word, Excel, PowerPoint et parfois Access), sous différentes versions (jusqu'à la version 95 pour la station HP). Il en est de même pour les utilitaires Internet courants : messagerie (Eudora) et navigateurs (Netscape).

Un logiciel très utilisé est EPI INFO, qui permet de traiter des données épidémiologiques (tableur, éditeur de texte...).

Enfin, selon les intérêts de chacun, on trouvera aussi : SPSS (statistiques), logiciels Internet autres (FTP, Telnet, Microsoft Explorer...), utilitaires divers (compression, audio, vidéo, antivirus, numérisation...), Microsoft FoxPro, 4D (bases de données), etc...

3. Le centre de documentation du C.CLIN Sud-Est : NosoBase

Comme nous l'avons vu auparavant, le C.CLIN Sud-Est doit remplir une mission documentaire, ce qui implique d'office l'existence d'un centre de documentation. Il est en place depuis la création du site Sud-Est. NosoBase, base de données bibliographiques sur les infections nosocomiales créée par le C.CLIN Sud-Est, constitue la principale (presque unique) activité du centre de documentation.

3. 1. Personnel

Nathalie LEROY est documentaliste et s'occupe (seule, à temps plein) du centre depuis décembre 1993. Sa formation a combinée documentation et médecine/biologie, elle connaît donc bien le domaine qu'elle côtoie.

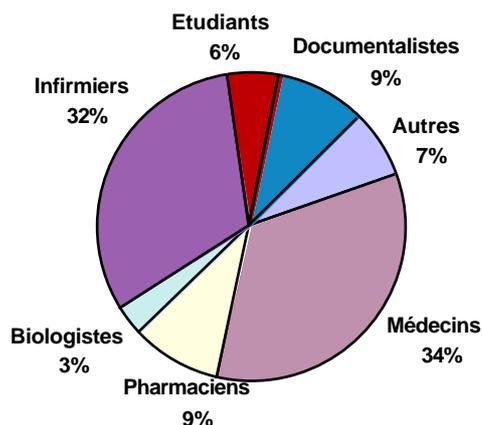
Les médecins du C.CLIN Sud-Est participent à l'indexation des documents de par leur spécialité.

En dehors de l'indexation, tout le travail repose sur les épaules de la documentaliste : abonnements, saisie, interrogation de la base, réponse aux questions, photocopies, classement, etc...

Récemment (septembre 1996), deux mi-temps ont été créés pour NosoBase : une première personne a été embauchée pour saisir les notices et participer aux réseaux de surveillance, une seconde pour le contrôle qualité de la base (contrôle du contenu principalement, mais aussi de la forme).

3. 2. Public

Le personnel du pavillon 1 M est souvent amené à consulter NosoBase (par l'intermédiaire de Nathalie LEROY : ils ne font pas la recherche seuls, en général). Dans l'ensemble, la visée de cette base s'adresse cependant à un public extérieur constitué de professionnels de la santé.



Graphique 1 : utilisateurs de NosoBase par catégories professionnelles

NosoBase s'est fait connaître auprès du public par voie publicitaire dans *HygieneS*, distribution de plaquettes de présentation auprès des présidents de CLIN, infirmiers et médecins hygiénistes, etc...

3. 3. Fonds documentaire

Le centre est abonné surtout "indirectement" aux revues : il reçoit les articles sélectionnés par son réseau de lecture (voir page 21; chaque membre du réseau possède personnellement son abonnement), et participe à la gestion commune des abonnements dans le pavillon 1 M.

Quelques monographies (guides professionnels, dictionnaires spécialisés, etc...) sont présents sur place, ainsi que 46 cassettes-vidéo, mais la majorité est un fonds de périodiques. Au total, la base informatique contient actuellement 2 385 notices (environ 100 nouvelles par mois), toutes les références étant issues de ce fonds. Les articles indexés sont photocopiés et conservés sur place, dans des dossiers classés.

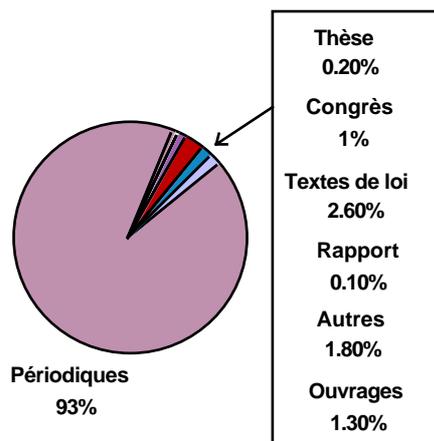
En ce qui concerne le contenu, NosoBase est spécialisée dans les écrits sur les infections nosocomiales en langue française. Cela doit lui permettre de conserver un attrait, une spécificité par rapport aux autres bases plus connues (Pascal, Medline, etc...). Bien évidemment, de nombreuses références sont en langue anglaise, indispensables dans tout domaine scientifique. Enfin, la majorité des documents sont des

articles

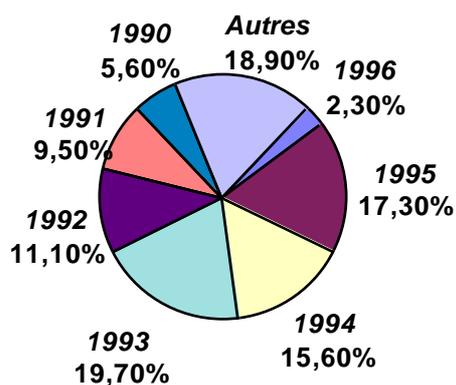
de

périodiques.

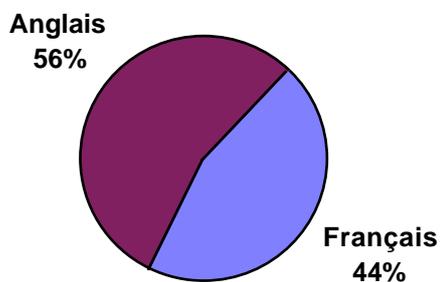
Voici plusieurs schémas sur les documents de NosoBase.



Graphique 2 : types de documents dans NosoBase



Graphique 3 : dates de publication des documents



Graphique 4 : langue des documents

3. 4. Activités

3. 4. 1. Sélection des articles et indexation

Onze centres de sélection des documents sont répartis dans toute la région Sud-Est, composés de médecins ou biologistes ayant chacun à charge une ou plusieurs revues de littérature française et internationale sur les infections nosocomiales.

Des médecins spécialistes du C.CLIN Sud-Est les indexent ensuite après tri pour intégration dans la base.

3. 4. 2. Saisie des notices

Jusqu'alors effectuée par Nathalie LEROY, ce travail devrait dès Septembre se répartir avec le mi-temps supplémentaire et la possibilité de saisir pour les autres C.CLIN (voir page 34).

3. 4. 3. Accueil du public

En raison de la disparité géographique du public, les interactions se font donc la plupart du temps par téléphone ou par courrier, même si les gens peuvent venir sur place.

3. 4. 4. Interrogation de NosoBase

3. 4. 4. 1. Accès indirect illimité

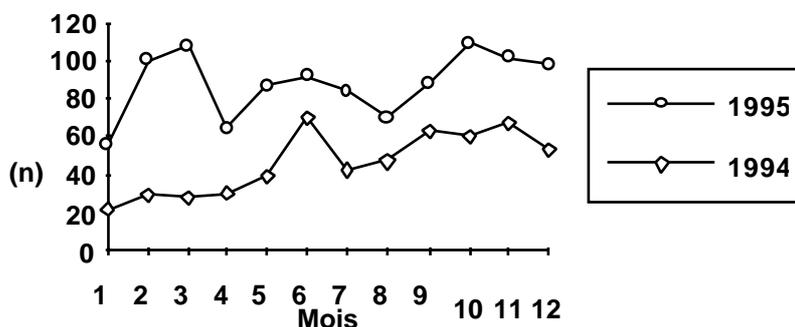
En dehors du Minitel (voir point suivant), l'accès à la base se fait sur un seul poste, celui de Nathalie LEROY, que les personnes extérieures doivent donc contacter (elle est, en général, la seule à répondre aux questions). Tout type de recherche documentaire (par sujet, auteur, année, etc...) est possible, sur l'ensemble des notices. Le résultat, s'il est quantitativement important, peut être envoyé par courrier, sur papier ou disquette, ou communiqué par téléphone.

3. 4. 4. 2. Accès direct limité

Le seul accès extérieur est le Minitel (36 17 BDSP). En effet, le C.CLIN Sud-Est est producteur pour la Banque de Données Santé Publique. Il s'agit d'une base de données bibliographiques générale sur la Santé en France, en langues anglaise et française en majorité, gérée par l'Ecole Nationale de la Santé Publique (ENSP). Plusieurs autres bases sont partenaires (BIRD, PASCAL, SAPHIR, TOXIBASE), et elle est alimentée par de nombreux producteurs (INIST, C.CLIN Sud-Est, centres de documentation d'hôpitaux, de santé publique, DRASS, observatoires de la Santé régionaux, etc...). L'INIST y tient une place dominante (environ 50 % des notices) : la BDSP retient ses notices en priorité dans le cas de doublons dans d'autres bases. Ceci est également valable pour le C.CLIN Sud-Est pour lequel, entre autre, toutes les notices ne sont pas acceptées d'emblée. C'est notamment le cas lorsqu'elles abordent un sujet trop précis (pas assez général en Santé Publique).

En conséquence, l'accès Minitel sur le 36 17 BDSP ne donne accès qu'à une partie de NosoBase (1 304 notices : environ 55 % de la base). L'interrogation correspond au mode minitel classique (interrogation par champs), et n'est pas très poussée.

3. 4. 4. 3. Statistiques des demandes au C.CLIN Sud-Est pour l'année 1995



Graphique 5 : nombres de demandes 1994/1995 au centre de documentation

3. 4. 5. Fourniture de documents

Les articles indexés sont photocopiés, placés à l'intérieur de leur fiche manuelle d'indexation (voir annexe) et rangés dans des dossiers. Le classement est numérique, en respectant l'ordre chronologique d'indexation.

La consultation se fait rarement sur place, mais plutôt par envoi de photocopies : 0,50 Frs la page au-delà de 50 pages, gratuit sinon. Le problème des droits de copies fait l'objet d'une réflexion dans le service, mais rien n'a été tranché en l'absence d'une institution de reversement des droits aux auteurs. Le Centre Français de la Copie vient apparemment d'obtenir (arrêté du 28 Juillet 1996 publié au B.O.) un statut légal, ce qui peut être une piste pour régler le problème.

Un service de prêt de cassettes-vidéo est aussi utilisable.

3. 4. 6. Produits documentaires

La revue bibliographique *NosoBase* est éditée trimestriellement et envoyée aux correspondants du C.CLIN Sud-Est (ou gratuitement sur demande). Elle est aussi intégrée à la revue *HygièneS*. Elle présente les articles récemment référencés sélectionnés par l'équipe (numéro de classification NosoBase, auteurs, titre, source), aux travers de plusieurs thèmes (*environnement, infections, matériel-produits, personnels, soins-prévention, surveillance (outils et méthodes)*). Ceux-ci doivent être prochainement modifiés.

Des dossiers thématiques sont réalisés, portant sur des sujets d'actualités (*Creutzfeldt-Jakob, Hépatite C, Risque professionnel et VIH...*).

Il est possible d'établir un profil documentaire et d'obtenir régulièrement le résultat de recherches bibliographiques sur papier ou disquette.

3. 4. 7. La base de données NosoBase

3. 4. 7. 1. Le logiciel documentaire

NosoBase fonctionnait auparavant sous le logiciel Mistral (société Bull). Le C.CLIN Sud-Est a pu bénéficier depuis décembre 1995, après un reformatage des données, du logiciel documentaire Basis plus acheté par les HCL. Trois autres centres de

documentation l'utilisent (Documentation et Pharmacie Centrales des HCL, Ecoles d'infirmières). Ce programme de la société ID France, en fonctionnement client-serveur, coûte très cher (environ 250 000 F pour dix utilisations simultanées en version client-serveur), et est en général utilisé par de grosses entreprises.

Le logiciel client fonctionne sous PC, bénéficie de l'interface de Windows 95, et possède les fonctionnalités suivantes : recherche par champs (opérateurs, tri des résultats, historique...), saisie, personnalisation de formulaires, index, thesaurus, exportation des données, etc...

La base et le moteur de recherche sont gérés physiquement sur le serveur UNIX de la Direction des Systèmes, de l'Information et de l'Informatique des HCL (DSII), avec les données d'autres bases.

Cette situation permet donc de bénéficier gratuitement d'un logiciel documentaire de bonne qualité, d'alléger le centre de documentation des tâches d'administration et de gestion de la base, mais pose un inconvénient majeur, à l'origine de ce stage (voir description page 27) : la connexion au serveur se fait sur la ligne des HCL, un serveur privé (Intranet) que les gestionnaires ne veulent pas ouvrir vers l'extérieur (Internet).

3. 4. 7. 2. Les notices

Les mots-clés sont normés d'après un *thesaurus NosoBase* : celui-ci, rénové régulièrement, a été développé en concertation avec les autres C.CLIN. Il est disponible sous trois formes (liste alphabétique, liste hiérarchique et liste hiérarchique détaillée) et n'est pas commercialisé.

On trouvera en annexe un exemplaire de la fiche d'indexation manuelle utilisée par les médecins pour rendre compte des articles. Les champs documentaires sont classiques, avec quelques particularités propres à la base (résumé obligatoire, titre original, candidat descripteur, commentaires...). On pourra se reporter au tableau page 38 pour une explication plus détaillée.

En ce qui concerne l'évolution de la base, Nathalie LEROY et Louis AYZAC ont commencé à réfléchir à une homogénéisation des données sous un codage de type SGML, pour pouvoir consulter les documents primaires après intégration dans la base par l'intermédiaire d'une station GED (OCR, archivage...).

3. 4. 8. Moyens informatiques

Nathalie LEROY dispose de deux ordinateurs PC : un 486 DX2-66 connecté au réseau des HCL pour interroger NosoBase, et un 486 SX-25 pour les autres travaux (accès au réseau Novell local, à Internet, bureautique). Le DX2 est suffisant pour Basis plus, mais l'autre poste est beaucoup trop lent et devrait être changé pour travailler sur Internet. Tous deux sont connectés à des imprimantes lasers.

La personne-ressource reste bien entendu Louis AYZAC.

3. 4. 9. Coopération inter-C.CLIN

Une première réunion inter-C.CLIN documentation a eue lieu à Pierre-Bénite le 10 Mars 1995. Des connaissances sur les fonds respectifs ont ainsi pu être apportées, les états d'informatisations, les produits, etc...

Les différents centres, chargés chacun de leurs régions, travaillaient de manière individuelle jusqu'alors et ont décidés d'unifier leur effort en créant une base commune, NosoBase. Le C.CLIN Sud-Est a donc été amené à jouer un rôle important.

4. Objectif du stage

Afin de permettre ce partage de ressources entre les C.CLIN, et ce sans trop de frais, le C.CLIN Sud-Est a décidé de mettre NosoBase à disposition par l'intermédiaire d'Internet. J'ai donc été "recruté" dans ce but.

II Mise à disposition de NosoBase sur Internet :

1. Cahier des charges

Voici l'étude préalable à la mise en place du projet. Il ne s'agit pas d'un cahier des charges élaboré selon les règles de l'art, mais d'une simple présentation.

1. 1. Faisabilité du projet

1. 1. 1. Les données du problème : réseau privé vs Internet

Comme nous l'avons déjà esquissé, le serveur Basis plus, avec les données de NosoBase, se trouve sur la machine UNIX des HCL, à l'intérieur d'un réseau privé. Le schéma suivant propose un aperçu de la situation.

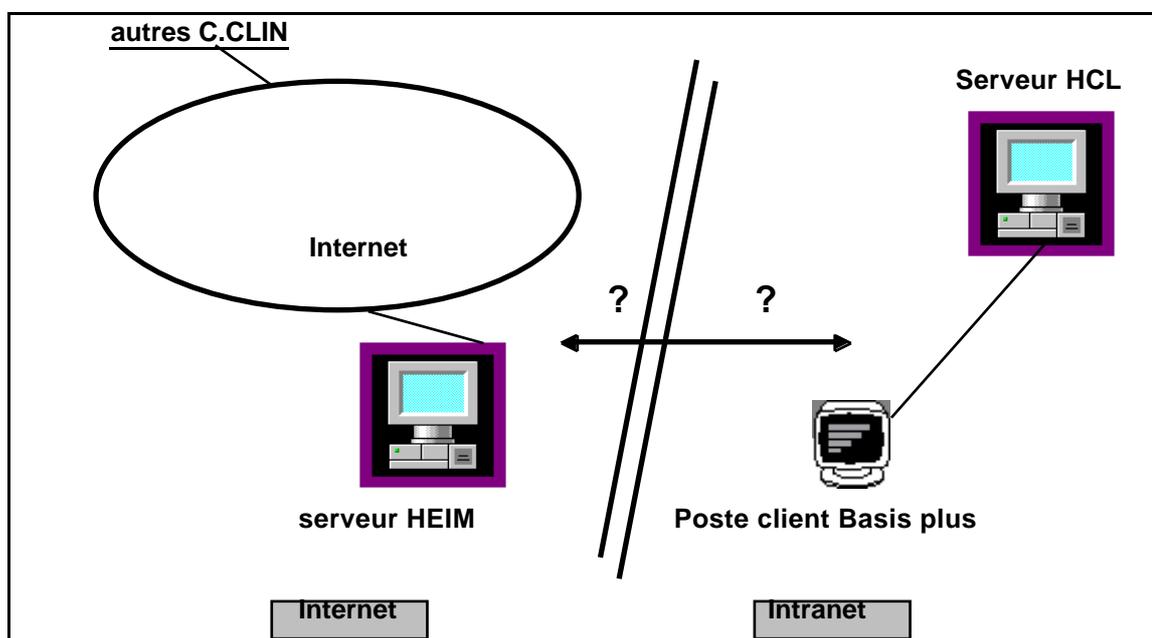


Figure 1 : le réseau privé des HCL

Nathalie LEROY doit donc disposer d'un poste dédié pour interroger NosoBase, connecté au seul réseau des HCL. Ceux-ci refusent que ce même poste soit connecté vers un autre serveur, ni Internet, ni le réseau local HEIM donc.

Cette solution, bien qu'évitant de gérer la base, est donc handicapante, même si les relations avec l'administratrice aux HCL sont bonnes.

Le poste client et le serveur qui concernent NosoBase appartiennent donc à un circuit fermé.

Cependant le C.CLIN Sud-Est reste propriétaire de ses notices, et les HCL sont tout à fait d'accord pour nous transmettre les données afin de les mettre à disposition par un autre moyen.

L'idée était donc de récupérer les données aux HCL, puis de les transporter vers une machine du C.CLIN Sud-Est ou de l'HEIM, ouverte vers Internet, et qui pourrait exploiter un moteur de recherche.

1. 1. 2. Coûts

Une étude préalable à mon stage avait été demandée à la DSII (service informatique des HCL), pour la consultation à distance de NosoBase. L'annonce des résultats avaient beaucoup traîné et la solution proposée était hors de prix :

Solution DSII	Connexion Numéris (3 h/j ; 20 j/mois)
	85 000 F/an/C.CLIN, dont :
Maintenance	21 500 F/an/C.CLIN
Consommation	5 327 F/mois/C.CLIN

Tableau 4 : la solution proposée par la DSII

Ceci comprend la connexion, sans devoir racheter le logiciel client Basis plus.

Acquérir un logiciel documentaire fonctionnant en client-serveur pour l'installer sur la machine de l'HEIM, autre solution, aurait été onéreux.

Si l'on ne considère pas le coût des heures passées (par le personnel du C.CLIN Sud-Est et moi-même) à élaborer l'ensemble du mécanisme, le transport vers Internet, sans étude préalable détaillée, ne devrait pas coûter beaucoup : le C.CLIN Sud-Est dispose d'un serveur, est connecté par l'intermédiaire de l'Université, et la plupart des moteurs de bases de données sur Internet sont gratuits.

1. 1. 3. Motivations et Conduite du projet

Une nouvelle rencontre inter-C.CLIN documentation était prévue à Pierre-Bénite le 4 Juin 1996. Les responsables de C.CLIN et les documentalistes étaient présents, ainsi que Louis AYZAC. Après un état des lieux de chaque centre, le projet de mise en réseau à été discuté. L'envie de partager ses ressources était général, et l'idée d'une base commune approuvée par tous.

La solution proposée par la DSII a été bien évidemment écartée, au profit d'Internet. Bien que peu connaissent le "réseau des réseaux", tous étaient assez attirés par cette solution, afin de pouvoir consulter NosoBase à distance et, certainement, pour ne pas rester en dehors du mouvement documentaire vers Internet. Il m'a semblé cependant ressentir une certaine méfiance vis-à-vis des réseaux, une appréhension, ou encore un doute sur leurs réels intérêts (l'aspect ludique, l'image de "l'internaute zappeur" primant parfois sur l'apport documentaire). Mais l'effet de groupe a masqué ces réticences.

Il a donc été décidé que le C.CLIN Sud-Est serait à la tête de ce projet, puisqu'il contrôlait la base, disposait d'une connexion à Internet, d'un serveur et... d'un stagiaire !

1. 1. 4. Conditions matérielles

1. 1. 4. 1. Connection

En ce qui concerne le C.CLIN Sud-Est, les conditions préalables au projet sont remplies : connexion à Internet, disposition de moyens matériels (données et serveur) et humains.

Les autres sites doivent être connectés, ce qui n'est à priori pas évident. Début Juin 1996, seulement deux sites avaient un accès au Réseau. En effet, les centres hospitaliers auxquels sont rattachés les C.CLIN n'offraient pas dans l'ensemble un branchement direct à Internet. Il fallait donc se connecter d'une manière personnelle, par modem. Or :

- aucun ne savait comment se connecter ni le prix à envisager. Il a donc fallu expliquer, renvoyer quelques descriptifs (configuration minimale, fournisseurs d'accès, ...) par la suite
- la demande d'un accès auprès des supérieurs hiérarchiques n'est pas simple à justifier (l'intérêt professionnel d'Internet est encore à prouver, sans compter la réticence dès qu'il s'agit de distribuer ses informations au public), et il faut ensuite compter avec la lenteur des administrations !

Actuellement, les démarches pour se connecter se trouvent à différents stades selon chaque C.CLIN.

1. 1. 4. 2. Logiciel

Un des avantages d'Internet est qu'il n'est pas nécessaire que les autres sites disposent d'un même logiciel documentaire : un simple "browser" (visualiseur Internet : Netscape ou Internet Explorer par exemple) suffira.

1. 2. Objectifs

Au travers d'Internet, 2 grandes possibilités devraient être offertes à termes :

- Consulter NosoBase

Il devrait être possible d'interroger le contenu de la base avec un outil disponible pour Internet, similaire à un logiciel documentaire. Les différents C.CLIN pourront ainsi disposer d'un fonds commun et mieux remplir leur mission d'information.

- Saisir des notices dans la base

Afin de participer pleinement au projet, tous les C.CLIN devraient pouvoir rentrer des données dans la base, le site Sud-Est n'étant plus le seul à faire ce travail. A court terme, un masque de saisie classique (Macros Word) sera réalisé pour mettre en oeuvre rapidement le circuit, puis une saisie sur Internet devrait pouvoir être envisagée.

1. 3. Répartition des taches entre les C.CLIN

C.CLIN Sud-Est
Chargé de la région Sud-Est
Direction du projet : mise en place et maintenance informatique
Saisie de notices
Collecte et vérification des données communes
Intégration des notices dans la base
Gestion des demandes de documents (recherche dans son fonds ou redirection vers d'autres C.CLIN)
Autres C.CLIN
Chargés de leur propre région
Saisie de notices après répartition des revues

Tableau 5 : répartition des tâches entre les C.CLIN

1. 4. Contenu

Lors de la réunion du 4 Juin, plusieurs points ont été soulevés. L'intérêt même de NosoBase, en tant que base commune nationale, a été discuté.

Il existe en effet plusieurs autres bases bibliographiques en médecine, dominantes sur le marché documentaire, souvent utilisées par les différents C.CLIN : Medline, Pascal, Current Contents...

La question de savoir si NosoBase se justifiait par rapport à ces dernières était donc légitime. Fallait-il poursuivre sur les mêmes assises ? Pourquoi continuer à indexer des documents présents dans Medline, par exemple ? Pourquoi ne pas reprendre dans NosoBase les notices d'autres serveurs ? La base doit-elle être exhaustive ?

La spécificité de NosoBase a ainsi pu être éclaircie. Une base spécialisée peut être nécessaire. Elle se justifie si elle apporte un complément :

- apport linguistique et culturel : une spécialisation dans les écrits concernant la France et la langue française, sans écarter la littérature anglaise, indispensable

- des revues spécialisées qu'on ne trouve pas ailleurs
- des articles pointus en infections nosocomiales, indexés finement (le thésaurus de Medline a été par exemple jugé trop général sur ce sujet, ni très satisfaisant pour les C.CLIN).

1. 5. Utilisateurs

A priori, les cinq C.CLIN doivent pouvoir saisir et consulter : il s'agit de l'objectif principal. L'ouverture du serveur pour tout public pourra être discutée par la suite.

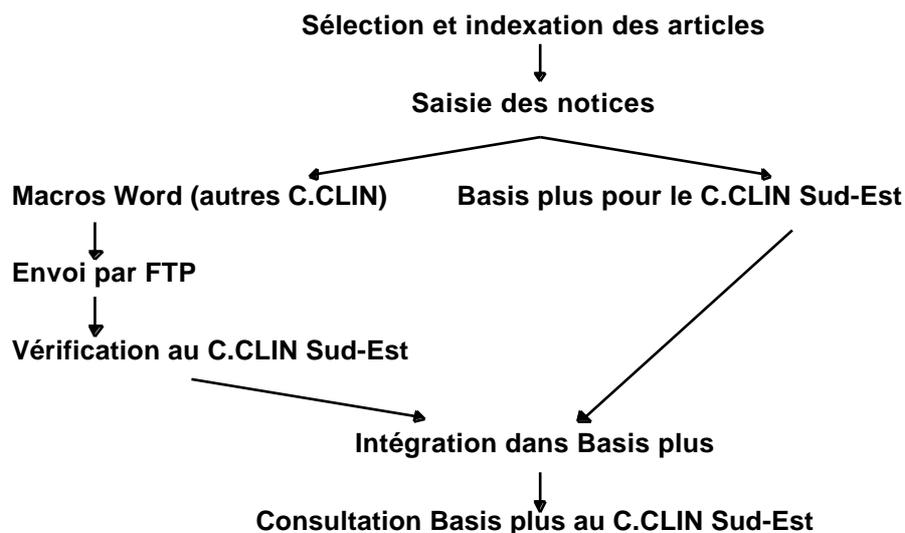
1. 6. Calendrier et étapes du travail

A compter du mois de Juin 1996, une phase pilote de six mois est lancée, pour apprécier la faisabilité du projet. La mise en place technique devra se faire avant quatre mois (durée du stage). Elle comporte :

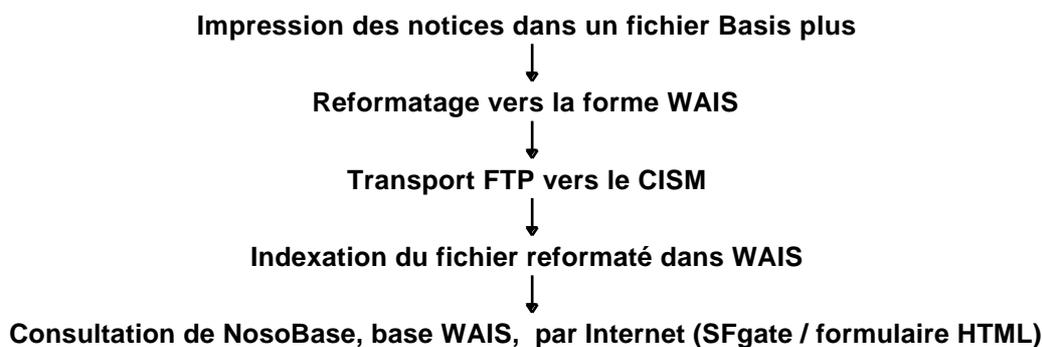
1. La création d'un programme de saisie des notices en local pour les autres C.CLIN
2. Le choix d'un moteur de base de données sur Internet
3. Le choix d'un serveur pour abriter la base
4. La mise en place des données et du moteur sur le serveur
5. La création d'une interface d'interrogation sur Internet
6. S'il reste du temps, la création d'un formulaire de saisie par Internet.

On trouvera en page suivante un schéma expliquant le circuit des données, de l'entrée sous Basis plus à la sortie Internet. Il aidera à mieux situer les points qui suivent.

1. Entrée des données sous Basis plus



2. Récupération des données de Basis plus pour WAIS



3. L'avenir : saisir par Internet

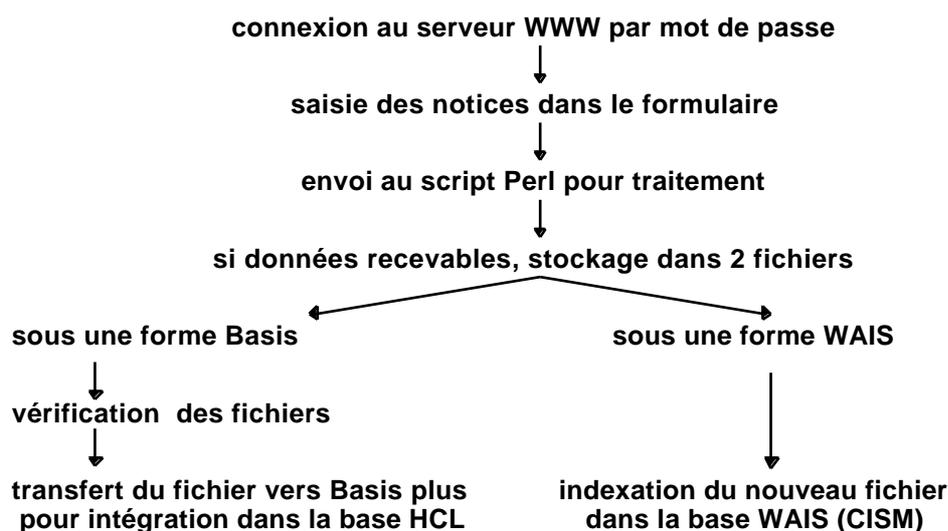


Tableau 6 : circuit des données Basis plus / Internet

2. Création d'un programme de saisie des notices en local

Avant de mettre en place NosoBase sur Internet, il importait de permettre la saisie des notices pour les autres C.CLIN.

2. 1. Circuit des données

Comme au C.CLIN Sud-Est, les personnes des autres sites doivent d'abord indexer les articles sur une fiche manuelle. Ensuite, elles devront utiliser le programme de saisie, installé sur leur poste en local, qui produira en sortie un fichier au format accepté par Basis plus. Celui-ci devra être envoyé au C.CLIN Sud-Est, par la poste ou par FTP sur le serveur HEIM, pour vérification par Nathalie LEROY et transmission aux HCL, qui l'intégreront en entrée de Basis plus.

En dehors de ce circuit des données, il ne s'agit donc que d'un programme de formatage des informations en syntaxe Basis plus.

2. 2. Choix du langage WordBasic

Lors de la réunion du 4 Juin, l'idée d'un masque de saisie à base de macros-commandes Microsoft Word avait été soumise.

2. 2. 1. Avantages

Le langage fourni avec Word me paraissait facilement abordable et fournissait quelques outils (boîtes de dialogues) facilitant l'interface avec l'utilisateur. Après plusieurs essais, il m'a semblé simple à mettre en oeuvre.

De plus, tous les C.CLIN disposait de ce traitement de texte.

Attrait principal, il permettait de manipuler directement du texte dans la fenêtre d'édition, à l'aide de nombreuses fonctions spécialisées, puisqu'il s'agit d'un traitement de texte.

Enfin, les données enregistrées pouvaient être visualisées par les C.CLIN avec ce logiciel.

Les macros Word peuvent se construire automatiquement, par enregistrement des actions de l'utilisateur dans Word. Ceci suffit en général pour de petites automatisations, mais il faut "mettre les mains dans le moteur" quand le traitement est plus poussé. Je m'en suis donc servi comme un langage de programmation structurée classique, en écrivant le code. Comme tous les logiciels Microsoft, une aide hypertexte est disponible par l'appui de la touche *F1* lorsque le curseur est placé sur une fonction : le développement s'apprend donc rapidement.

2. 2. 2. Inconvénients

Word Basic (le langage des macros-commandes) n'est pas un "véritable" logiciel de programmation, puisqu'il s'appuie sur un traitement de texte. Bien qu'il fournisse des outils classiques (manipulations de fichiers, entrée/sortie, structures conditionnelles, boucles, tableaux...), il n'est pas fait pour construire de gros programmes, mais pour rendre quelques services ici et là.

Le tout fonctionne alors que Word est chargé en mémoire (il n'est donc pas très rapide pour de longs traitements), l'interpréteur n'est pas un exemple de vitesse, et les fonctions restent sommaires. On ne peut ainsi pas créer de fichiers exécutables (le programme est donc dépendant de la version Word utilisée, et devra être retranscrit par l'avenir), il n'y a pas d'outils de débogage, ...

Cette solution a été facile à développée, elle fonctionne, mais une reprise avec un langage plus performant serait certainement souhaitable par la suite. N'oublions pas non plus qu'une éventualité de saisie par Internet (voir page 77) a été envisagée, ce qui rendrait théoriquement inutile un tel masque de saisie (sauf pour les C.CLIN qui ne disposeraient pas encore d'une connexion).

2. 3. Format à obtenir

NosoBase sous Basis plus est gérée par la DSII des HCL. Plusieurs contacts avec l'administratrice de la base et l'aide d'anciens documents (issus de la reprise de NosoBase sous Mistral, ancien logiciel, vers Basis plus) m'ont permis de définir le format en sortie.

Le tableau page suivante présente les champs de NosoBase, tels qu'ils apparaissent dans la fiche manuelle d'indexation (voir annexe), avec leur intitulé exact en code Basis plus et les contraintes quant au contenu.

Intitulé sur la fiche manuelle d'indexation	Code Basis plus	Obligatoire	Longueur maximale (caractères)	Particularités - Exemple
n° d'indexation	#REF	oui	10	Numéro interne de classification. Se remplit automatiquement exemple : 0000003001
Type de document	#TYPE	oui	1	Liste d'autorité : <i>O = Ouvrage ; P = Périodique ; T = Thèse ou mémoire ; R = Rapport ; C = Congrès ; L = Législation ; A = Autres</i>
Titre 1	#TITRE_FR	oui	300	Titre français
Titre 2	#TITRE_OR	non	300	Titre en langue originale
Auteur(s)	#AUTEUR	oui	900	La forme générique est : nom en majuscules + espace + initiale du prénom en majuscule + point
Adresse	#ADR_AUT	non	200	Adresse de l'auteur
Source	#SOURCE	oui pour un périodique ou un ouvrage	270	Périodique ou ouvrage d'où est tirée la partie mentionnée dans <i>Titre 1</i> . Les initiales des mots de périodiques (sauf les articles) doivent être en majuscules
Editeur	#EDITEUR	oui pour un ouvrage	100	
Date de publication	#DATE	oui	10	Format obligatoire : AAAA/MM/JJ, AAAA/MM ou AAAA (avec AAAA = année sur 4 chiffres, MM = mois sur 2 chiffres, JJ = jour sur deux chiffres)
ISSN	#ISSN	oui pour un périodique	20	
ISBN	#ISBN	oui pour un ouvrage	20	
Tome	#TOME	non	10	
Volume	#VOLUME	non	10	
Numéro	#NUMERO	non	20	
Page(s)	#PAGES	non	15	Forme générique : numéro première page + tiret + numéro

				dernière page, ou nombre de pages
Langue	#LANGUE	oui	24	Liste d'autorité : <i>Français = FRE ; Anglais = ENG ; Allemand = GER ; Espagnol = SPA ; Italien = ITA ; Russe = RUS ; Japonais = JPN ; Autre = AUT</i>
Pays	#PAYS	non	20	Liste d'autorité : <i>France = FRA ; Grande-Bretagne = GBR ; Allemagne = DEU ; Espagne = ESP ; Italie = ITA ; Russie = RUS ; Japon = JPN ; Etats-Unis = USA ; Autres = AUT</i>
Bibliographie	#BIBLIO	non	10	Nombre de références mentionnées dans la bibliographie du document indexé
Genre de document	#GENRE	oui	1	Liste d'autorité : <i>1 = Editorial, Opinion, Interview ; 2 = Revue générale, synthèse ; 3 = Compte-rendu de Recherche ; 4 = Présentation de Matériel et produit ; 5 = Recommandations et protocoles ; 6 = Autres</i>
Contenu	#MOTSCLES	oui	1200	En Majuscules - Mots-clés du <i>Thesaurus NOSOBASE</i>
Contexte	#CONTEXTE	non	500	Indexation libre (mots n'appartenant pas au thesaurus : candidats descripteurs)
Résumé 1	#RESUME_FR	oui	3 000	Résumé en français
Résumé 2	#RESUME_OR	non	3 000	Résumé en langue originale (à recopier à partir de l'article)
Commentaire (si nécessaire)	#COMMENT	non	300	Il s'agit des trois rubriques de la fiche d'indexation (sous <i>Commentaire (si nécessaire) : valeur scientifique - nouveauté de l'apport - intérêt pour la pratique</i>)
-----	#INDEXEUR	oui	20	Ce champ, absent de la fiche manuelle d'indexation, contient le nom du C.CLIN indexeur de la notice. Jusqu'à maintenant, il s'agissait toujours du site Sud-Est, il a donc été ajouté dans le programme pour les autres C.CLIN. Il se remplit automatiquement.

Tableau 7 : champs Basis plus

Le codage s'effectue de la façon suivante :

- début de notice : @A NOSO
Il sert aussi de séparateur entre les notices
- codage des champs : #nom Basis du champ=valeur
un champ par ligne, suivi donc d'un saut de paragraphe.
Pas de saut de ligne à l'intérieur d'un champ

Le tout est contenu dans un ou plusieurs fichiers au format texte (texte windows ANSI)

1. 1. Le programme en macros-commandes Word

Pour pouvoir exécuter les macros-commandes, il faut disposer de Microsoft Word, bien sûr, et les copier dans le répertoire adéquat : une macro-commande est attachée à un modèle de document (les fichiers *.dot*), il suffit donc de placer celui-ci dans le répertoire réservé aux modèles de Word. Une option *Exécuter* est alors disponible dans le menu *Outils, Macro*, mais le tout a été automatisé afin que le programme puisse être lancé directement de la barre d'outils Word.

Les macros ont été développées sous PC (une version pour Mac, plus pratique pour un des sites, est en cours d'étude), pour les versions 2 et 6 de Word (seules quelques commandes changent). Le modèle de document à récupérer s'appelle *Nosobase.dot*. Il contient les macros suivantes :

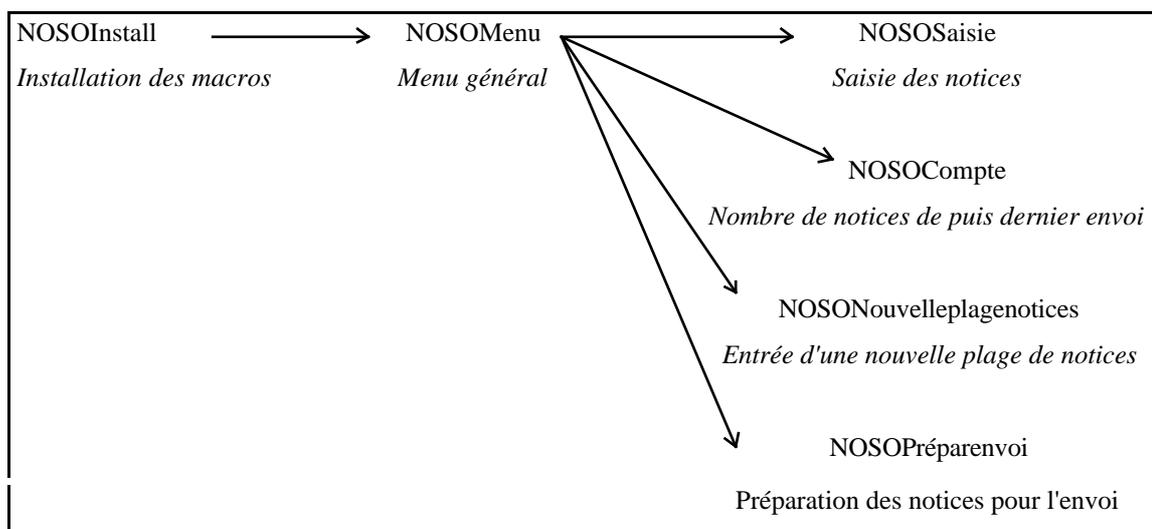


Figure 2 : programme de saisie en local

1. 1. 1. NOSOInstall : macro à exécuter uniquement à la première installation

- Copie les macros dans le modèle *normal.dot*
- Affectation d'un icône  sur la barre d'outils et d'un sous-menu *NOSObase* dans le menu *Fichier*, tous deux dédiés à la macro NOSOMenu.
- Création des répertoires utiles au programme et d'un fichier de configuration (contient le numéro de la notice suivante à saisir, le nombre de notices saisies depuis la dernière préparation pour l'envoi, la borne supérieure de la plage de notices autorisée pour le C.CLIN, etc...)
- Demande à l'utilisateur son identité de C.CLIN et la saisie de sa plage de notices (contrôle si borne supérieure > borne inférieure) : celle-ci est fournie par le C.CLIN Sud-Est et permet de répartir entre les sites les numéros d'indexation à donner aux notices saisies, ce afin d'éviter des doublons dans la base commune. Les interactions avec l'utilisateur se font par l'intermédiaire de boîtes de dialogues classiques (message seulement, demande de confirmation ou d'entrée de valeur), comme celle ci-dessous :

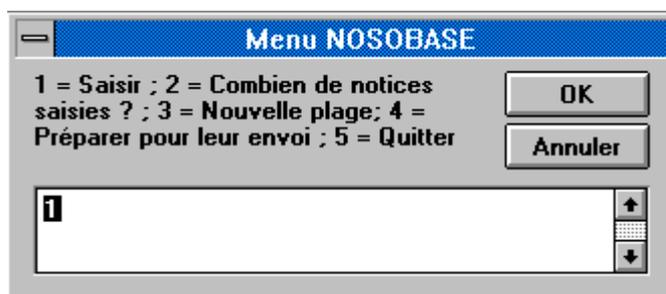


Image 1 : boîte de dialogue du masque de saisie

- Lance enfin la macro NOSOMenu.

1. 1. 2. NOSOMenu : menu général

Il s'agit du menu général disponible une fois le programme installé. Il propose plusieurs options à l'utilisateur, et peut donc être appelé à partir du bouton  de la barre d'outils ou du menu *Fichier*, *NOSObase*. Voici les choix proposés :

1. Saisie des notices ↩ance la macro NOSOSaisie
2. Donne le nombre de notices saisies depuis la dernière "préparation à l'envoi" ↩OSOCompte
3. Pour rentrer une nouvelle plage de notices ↩OSONouvelleplagenotices
4. Prépare les notices saisies pour un envoi ↩OSOPrépareenvoi
5. Quitte le programme

1. 1. 3. NOSOSaisie : saisie des notices

La personne doit se reporter à la fiche manuelle d'indexation sur laquelle se trouveront toutes les informations à saisir.

Les valeurs de tous les champs sont demandées, parfois remplis automatiquement (le champ *INDEXEUR* par exemple).

Certaines vérifications sont effectuées (si la valeur est correcte dans le cas d'une norme à respecter, le champ *DATE* par exemple ; ou si elle est obligatoire). Le remplissage de certains champs peut conditionner l'obligation d'autres (s'il s'agit d'un périodique, alors le champ *SOURCE* devient obligatoire).

En fonction des réponses, une notice se remplit en arrière-plan des boîtes de dialogues, sur une feuille Word classique :

- avant la saisie, le numéro de référence de la notice est incrémenté et s'inscrit automatiquement après le séparateur de notice (en début de saisie, un contrôle est d'abord effectué pour savoir si la plage de notices autorisée n'est pas dépassée : voir NOSOInstall)
- après vérification pour chaque champ, l'intitulé Basis plus s'inscrit, suivi de la valeur entrée. Un champ vide n'a pas lieu d'être et n'est donc pas inséré.

En fin de notice, il est possible de continuer la saisie tant que la plage n'est pas dépassée : le tout s'inscrit à la suite sur une même feuille.

En fin de saisie, le tout est enregistré sous un nom de fichier normé (numéro de la dernière notice saisie + code du C.CLIN indexeur) sous différents formats (Word pour une relecture, copie de sauvegarde et texte pour Basis plus) et dans différents répertoires

(certains serviront à l'envoi, d'autres à la consultation des notices saisies...). Enfin, les valeurs du fichier de configuration sont remises à jour.

1. 1. 4. NOSOCompte

Affiche simplement le nombre de notices saisies depuis la dernière "préparation à l'envoi" (voir NOSOPrépareenvoi), stocké dans le fichier de configuration.

1. 1. 5. NOSONouvelleplagenotices

Demande les bornes (inférieure/supérieure) de la nouvelle plage de notices (voir NOSOInstall), dans le cas où celle-ci serait dépassée, et les enregistre dans le fichier de configuration. Il faut auparavant contacter le C.CLIN Sud-Est et demander une nouvelle plage.

1. 1. 6. NOSOPrépareenvoi

Cette macro permet de préparer les notices afin qu'elles soient envoyées au C.CLIN Sud-Est :

- place les fichiers au format texte dans un repertoire *AENVOYER*
- initialise correctement le fichier de configuration et les répertoires de stockage (sauf un qui contient les notices saisies depuis la première installation du programme, pour consultations ultérieures).

Il ne reste plus au C.CLIN indexeur qu'à récupérer les notices et nous les envoyer par La Poste ou par FTP.

1. 1. 7. FTP

Avec la disquette d'installation, un programme FTP (File Transfert Protocol : transfert des fichiers) gratuit (*freeware*) a été fourni, pour envoyer les fichiers par Internet. Il s'agit de *Windows Sockets FTP Client (WS_FTP)*, récupéré sur le Réseau.

1. 2. Matériel fourni

Une disquette a été distribuée aux quatre autres C.CLIN, contenant :

- macros pour Word 2 et Word 6

- programme FTP
- document Word (manuel d'utilisation)

Un manuel papier a été élaboré et distribué : installation, utilisation, envoi par FTP, exemples...

1. 3. Test

La saisie des notices par les autres C.CLIN nécessitent qu'ils se répartissent auparavant les revues à indexer, pour éviter des travaux parallèles identiques qui seraient inutiles. Les démarches sont en cours, en Septembre 1996, et aucune notice n'a encore été produite de cette manière.

La disquette et le manuel ont été envoyés et, passées les premières difficultés de compréhension, le programme semble avoir été bien accepté : quelques remarques positives nous sont ainsi parvenues.

Dès que le système sera en place, il faudra donc tester rapidement la chaîne totale :

saisie des notices → envoi au C.CLIN Sud-Est → transfert vers Basis plus (DSII)

C'est pourquoi il a été demandé d'envoyer les fichiers au bout de 10 notices saisies, pour corriger d'éventuelles erreurs.

Enfin, un dossier de programmation a été réalisé pour Louis AYZAC et Nathalie LEROY, afin de reprendre le programme quand je ne serai plus là.

2. NosoBase sur Internet

2. 1. Choix du moteur de recherche et du serveur

Un choix important était ensuite celui du coeur du système : le moteur de recherche sur Internet. Il fallait aussi décider où placer les données, c'est-à-dire trouver un serveur.

2. 1. 1. Le moteur de recherche : WAIS

Le langage HTML classique permet d'écrire des liens, de monter un serveur hypermédia, mais pas d'interroger des bases de données. Il faut rajouter aux pages du serveur un moteur de recherche qui ira questionner la base et ramener les données pertinentes.

De petits outils ont bien été développés sur Internet, mais rarement pour gérer plusieurs milliers de documents.

Monsieur LARDY m'avait aussi parlé de *Glimpse*, mais il ne semblait pas encore beaucoup utilisé.

Notons enfin qu'un article d'*Archimag* (voir bibliographie) signalait un complément Internet pour Basis plus de la société ID France nommé *Basis server* : cette possibilité aurait été idéale, mais payante, et les HCL ne voulant pas ouvrir leur réseau, cela nous aurait obligé à racheter Basis plus avec sa solution serveur : impensable !.

En parcourant Internet ou en feuilletant de la littérature spécialisée préalablement au stage, le seul moteur du domaine public satisfaisant était WAIS, par ailleurs largement utilisé (voir description plus loin).

Plusieurs versions des logiciels WAIS se sont retrouvés sur le marché, gratuites pour la plupart, et il aurait donc fallu faire un choix. En parcourant les différents documents sur Internet, FreeWAIS-sf, de l'Université de Dortmund (Allemagne), nous a semblé d'emblée le plus performant. Parmi les plus utilisés, il intégrait des possibilités absentes des premiers logiciels WAIS, dont l'utilisation de champs dans les documents. C'est le choix du serveur qui nous a de plus orienté vers cette solution.

2. 1. 2. Le serveur du Centre d'Information Scientifique et Médicale (CISM) de l'Université Lyon 1

WAIS est dévoué en majorité au monde Unix. Une liste des plateformes disponible pour FreeWAIS-sf existait bien sur le serveur de Dortmund, mais confirmait cette donnée : pas de version pour Novell Netware. A part monter un serveur Linux (Unix sous d'autres plateformes, dont MS-DOS), ce qui n'enchantaient franchement pas

Louis AYZAC, nous ne pouvions donc pas abriter une base WAIS sur le serveur Novell de l'HEIM.

Louis AYZAC avait par ailleurs pris plusieurs contacts avec des personnes concernées par WAIS. Il est enseignant (Biostatistiques-Informatique en deuxième et troisième année de Médecine ; Maîtrise de Biologie Humaine, Licence et Maîtrise de Sciences Sanitaires et Sociales) et responsable réseau à l'Université Lyon-Sud de Médecine. Cette faculté, située à 200 mètres du pavillon 1 M, est rattachée à l'Université Lyon 1. A ce titre, il s'est avéré que nous pouvions faire héberger notre base sur les machines du CISM, Centre d'Informatique Scientifique et Médicale de l'Université Lyon 1 (ancien Centre de Calcul).

Colette Maddalena est administratrice WAIS au CISM. Elle a déjà monté plusieurs autres bases (dont celle de la bibliothèque de l'ENSSIB, de l'INSA...), et a accepté d'abriter, de construire et de maintenir NosoBase sur leur serveur. Elle est très serviable, efficace, et nous a donc beaucoup aidée.

Leur logiciel est par ailleurs FreeWAIS-sf dans sa version 2.0.

Nous présentons d'abord WAIS dans son ensemble, puis la version utilisée par le CISM.

2. 2. WAIS

Je n'ai pas eu personnellement à implémenter les programmes WAIS sur les machines du CISM, puisque c'était déjà fait, ni à construire les réputés fastidieux fichiers descriptifs d'une base. Je me suis consacré à l'interfaçage avec des clients WWW et au reformatage des données. C'est un (petit) regret, puisque je n'ai pas eu à mettre sur pied une "chaîne WAIS" complète du début à la fin, mais l'objectif était de mettre rapidement NosoBase sur Internet. Nous avons donc profité de l'expérience du CISM, à qui nous avons laissé cette tâche.

Cependant, la connaissance "théorique" de Wais et de FreeWAIS-sf m'a paru indispensable lors de mes contacts avec l'administratrice du CISM, pour savoir ce qui était possible, comment certaines modifications pouvaient être réalisées, etc...

Il me semble maintenant que l'installation d'une base WAIS est tout à fait réalisable, avec de la documentation, les programmes et machines adéquats, une connaissance du monde Unix, des serveurs et des données documentaires, et beaucoup d'huile de coude !

Nous présentons donc WAIS, dans cette optique : les connaissances nécessaires pour monter une base à l'aide d'un spécialiste

2. 2. 1. Le passé

Lors des entretiens précédants le stage, nous avons pensé à WAIS pour mettre NosoBase sur Internet. Je m'y suis donc intéressé avant le mois de Juin, et j'ai commencé par des articles de présentation générale. WAIS existe depuis le début des années 1990, et il est vrai que certains écrits qui datent parfois de 1995 (c'est encore récent !) abordent une situation qui n'a plus lieu d'être. Passé l'enthousiasme des débuts de WAIS, soutenu par le projet CNIDR (Clearinghouse for Networked Information Discovery and Retrieval, de la NSF américaine), qui aurait dû assurer son développement, les choses ont changées. Nous décrivons donc d'abord le commencement de WAIS, avant de nous intéresser à la situation actuelle.

2. 2. 1. 1. Présentation

Wais (Wide Area Information Servers) est un ensemble de logiciels développés par les sociétés Dow Jones, Thinking Machines Corporation et Apple Computers. Ces outils d'indexation et de recherche de documents sur les réseaux TCP permettent de :

- constituer des index à partir de mots contenus dans des fichiers de formats divers (texte Ascii, Postscript, fichiers HTML, Mime, images, sons...)
- interroger ces index en utilisant une syntaxe d'interrogation similaire aux logiciels documentaires
- rapatrier les documents sélectionnés et les visualiser.

Ces programmes fonctionnent sur le modèle client-serveur et s'appuient sur le protocole ANSI Z39.50.

2. 2. 1. 2. La norme Z39.50

Il s'agit d'un protocole conçu pour uniformiser les dialogues entre un client et un serveur dans le cadre d'applications documentaires, s'inscrivant dans la philosophie du modèle OSI de l'ISO comme un protocole d'application, défini par l'ANSI (organisme de standardisation aux USA). Il s'appuie sur des dialogues en mode connecté et fournit des services pour l'initialisation de la communication client-serveur, le déclenchement de recherches, la récupération des résultats et la terminaison de la communication.

Deux versions de cette norme sont supportées par les logiciels WAIS : Z39.50 88 et Z39.50 92, la plus récente.

2. 2. 1. 3. Les logiciels Wais

Plusieurs versions ont été développées, dont :

- WAIS 8-b5.1 créée par Thinking Machines Corporation
- des variantes de ce dernier (wais-b5.2), développées par l'INRIA et utilisées surtout en France
- FreeWAIS 0.3, mise au point par le CNIDR
- FreeWAIS-sf développée par l'université de Dortmund (Allemagne), qui semble être actuellement la plus utilisée.

Toutes ces versions sont gratuites, seul un programme a été commercialisé par WAIS Inc., surtout utilisé aux USA (il ne semble plus disponible).

Toutes les versions serveurs se destinent principalement à une implémentation sous Unix, bien que d'autres supports soient possibles. Les logiciels clients existent pour tout type de poste.

Le "package" Wais se compose de plusieurs logiciels :

- un programme d'indexation permettant de construire sur le serveur un ou plusieurs index des documents interrogeables : *waisindex*
- un programme client qui permet d'interroger les index sur des serveurs : *waissearch*
- un programme serveur qui consulte les index de la base et peut répondre aux demandes des clients : *waisserver*.

2. 2. 1. 4. Les bases WAIS

Tout peut être indexé sous WAIS, mais la plupart des bases contiennent des données bibliographiques. Les fonds documentaires étaient au départ majoritairement issus des sciences "dures" (mathématiques, physique...), de laboratoires, universités, écoles normales supérieures, INRIA... On trouve peu de texte intégral, il s'agit surtout de notices.

2. 2. 1. 5. Interrogation des bases

Pour avoir accès aux bases, il fallait auparavant disposer d'un logiciels client utilisant *waissearch*, disponible sous différents plateformes (exemple : *Winwais* sous Windows).

Il était possible d'interroger une ou plusieurs bases en rentrant sa question dans une syntaxe particulière (nom des champs, opérateurs et mots-clés). Le serveur recevait la requête, fouillait dans le ou les index de la base (construits par *waisindex*) et retournait les lignes d'en-tête des documents (ligne descriptive de la référence) correspondants, triés par ordre de pertinence (système de pondération de la question mise en parallèle avec le poids de chaque terme dans les documents) : il suffisait alors de sélectionner une ou plusieurs de ces lignes pour avoir accès au texte intégral. Lors d'une question suivante, on pouvait aussi bénéficier du mécanisme de *relevance feedback* : cela permettait d'affiner la question en lui ajoutant les contenus des documents jugés particulièrement pertinents par l'utilisateur.

Pour avoir accès à un serveur, il fallait connaître ses coordonnées : adresse IP du serveur, nom et port TCP (souvent 210) de la base. Celui-ci pouvait être privé (accès par mot de passe), semi-public (disponible mais adresse non communiquée), publique au niveau national ou à l'échelle internationale.

A cet effet, plusieurs *directory of servers* étaient maintenus : il s'agissait d'annuaires des bases WAIS signalées par leur administrateur, interrogeables comme tout autre base. Au lieu de communiquer des notices bibliographiques, ces serveurs particuliers retournaient des documents descriptifs des bases enregistrées (fichiers *.src* : coordonnées de la base, nom des champs, type de données, contacts...). A ce titre, ils

étaient bien pratiques, comme les robots disponibles actuellement sur Internet (Altavista, Yahoo, Veronica, Archie...).

Assez nombreux, on pouvait trouver celui de Thinking Machines Corporation, du CNIDR, de la NASA et, en France, le *directory of servers* de l'INRIA (toujours existant : voir *Bookmark*) pour les bases françaises.

Il est encore possible d'utiliser un client spécifique WAIS pour accéder aux bases, mais les choses ont vraiment changées en ce qui concerne ces "annuaires" : ils n'existent plus.

2. 2. 2. Le présent

Cette séparation passé/présent s'explique par la question que nous avons dû nous poser en se basant sur un état des lieux dépassé : fallait-il adopter WAIS ?

En effet, après avoir bien parcouru différents articles, j'ai commencé à essayer d'interroger des bases avant le stage, à partir du client *Winwais* pour Windows à l'ENSSIB. L'aide disponible dans le logiciel est tout à fait correcte et suffit pour commencer. De plus, une longue liste d'adresses de serveurs est fournie.

Cependant, il était rare que j'arrive à me connecter à une base, encore moins à un des *directory of servers*.

2. 2. 2. 1. WAIS existe toujours

Après de nombreux tâtonnements, j'ai finalement appris (l'information ne se trouvait nul part officiellement) que de nombreuses bases n'existaient plus (les aléas d'Internet), et que les *directory of servers* étaient presque tous abandonnés. Quelques sites ont parfois conservé d'anciennes listes de serveurs, rarement mises à jour. Fallait-il donc investir du temps dans une solution en perdition ?

Cette interrogation a persisté jusque quinze jours après le début du stage. Chose étonnante, des messages passés dans des listes ou groupes spécialisés, pourtant assez provocateurs (avec le sujet : "Wais est-il mort ?"), n'ont rencontré que peu d'écho, assez cependant pour comprendre la situation :

- les initiatives officielles internationales d'entreprises qui souhaitent appuyer le protocole WAIS, comme Thinking Machine Corporation ou le CNIDR, ne sont plus très retentissantes ou ont disparues.
- de même pour les *directory of servers*.

L'aspect de WAIS semble s'être modifié : il existe toujours beaucoup de bases, mais l'interrogation ne se fait plus par un client spécifique, plutôt au travers de serveurs HTTP classiques. Derrière les nombreux formulaires que l'on peut maintenant trouver sur des pages HTML se cachent peut-être des bases WAIS : l'interfaçage s'effectue maintenant par des passerelles WWW-WAIS.

Il n'y a donc plus intérêt à alimenter un monde WAIS parallèle au WWW classique (HTTP), avec ses clients, ses annuaires, mais le tout se fond dans le "web" actuel avec d'autres moteurs.

WAIS reste cependant une solution très prisée pour disposer gratuitement d'un moteur de recherche.

Notons enfin qu'il n'est pas évident d'obtenir une information correctement mise à jour, exhaustive et correcte à propos d'Internet, étant donné la rapidité de changement et le "gigantisme" du système. Mon point de vue sur WAIS ne correspond peut-être pas à l'exacte vérité, mais rien n'est venu lors du stage me prouver le contraire.

2. 3. FreeWAIS-sf

Il s'agit donc du logiciel WAIS disponible sur le serveur du CISM qui abritera NosoBase.

Développée par l'Université allemande de Dortmund, cette version gratuite (*Free*) permet d'exploiter des documents avec champs (*-sf* pour *search fields*) en construisant des fichiers inverses spécifiques. Actuellement, la dernière mise à jour est la 2.1.

2. 3. 1. Installation

Sur tout système Unix, le système se compose d'un démon *waisserver* et d'un indexeur *waisindex*. Les fichiers doivent être installés dans le répertoire adéquat, le

système configuré (nous n'entrons pas dans les détails), puis il faut indexer les fichiers et mettre en place le démon. Une grande attention devait auparavant être apportée à l'élaboration du fichier *.src*, celui-ci décrivant la base à l'interrogation d'un *directory of server* (à condition de la signaler au niveau national ou international).

2. 3. 2. Le logiciel d'indexation

Il s'agit d'un programme qui va parcourir un ou plusieurs fichiers de documents sur le serveur et constituer un index des mots d'au moins deux caractères. Il faut lui préciser (dans un fichier *nom de la base.fmt*) :

- le format des documents (texte, gif, HTML...)
- les règles de découpages des fichiers en documents (quel est le séparateur entre les document)
- la structure des champs : un ensemble de caractères réservés (*regular expressions*) comme en Unix permet de coder le début et la fin d'un champ, la présence d'une suite de caractères majuscules ou minuscules, de ligne vide...
- la façon dont sera constituée la *headerline*, ligne d'en-tête descriptive du document qui sera rapatriée avec d'autres en résultat d'une interrogation.

Cette partie a la réputation d'être fastidieuse. Elle nécessite donc l'indexation de documents correctement structurés.

L'exécution de la commande *waisindex* crée un certain nombre de fichiers, dont un index global (indexation de l'ensemble des fichiers) et des index locaux (un index par champ). Ces index sont des fichiers inverses qui contiennent une entrée par mot, avec :

- les références des documents qui le contiennent
- sa position dans le document
- le poids du terme dans le document (qui sera utilisé pour pondérer les résultats d'une requête)
- la position de la première occurrence du mot dans les documents, et des suivantes s'il y a lieu.

La taille de la base finale peut être multipliée de 0,3 à 3 fois par rapport aux fichiers originaux.

Lors d'une mise à jour, il faudra réindexer l'ensemble des documents.

2. 3. 3. Syntaxe d'interrogation

A l'origine, les questions WAIS consistaient en un ou plusieurs mots séparés par un espace.

FreeWAIS-sf, avec d'autres, a rajoutée l'utilisation d'opérateurs, la mention des champs à interroger, etc...

2. 3. 3. 1. Champs :

Lorsque celui-ci n'est pas précisé, l'interrogation se fera sur l'index global, c'est-à-dire l'ensemble des documents.

Il est possible de rechercher sur un champ seulement :

nom WAIS du champ=valeur

Exemple : *AUTE=Flaubert*, *AUTE* étant l'intitulé exact du champ auteur dans les documents indexés sous WAIS).

2. 3. 3. 2. Opérateurs

On trouvera les opérateurs suivants :

- les "classiques" AND, OR, NOT, qui peuvent être utilisés pour connecter les valeurs d'un même champ ou plusieurs champs entre eux (ex : *AUTE=Flaubert AND TITRE=Bovary ; Flaubert ET RESUME=roman*). Il est possible de déterminer un opérateur par défaut, parmi ET ou OU, lorsque de simples mots sont utilisés sans être reliés par des opérateurs (interrogation en "langage naturel")
- La troncature à droite : * (ex : *MOTCLE=chirurg** ← chirurgie(s), chirurgien(s), surgical...)
- expression littérale : " " (ex : "*information retrieval*")
- opérateurs pour champs numériques (il existe un format texte, numérique, date...): ==, >, <, >=, <=

- opérateur de proximité :
 - phonétique : soundex (ex : *AUTE=soundex salatan*, permettra de trouver "salton"). Adapté certainement à l'anglais uniquement
 - proximité classique : w/ (ex : *énergie w/10 nucléaire* recherchera les phrases qui contiennent ces deux termes séparés au maximum par dix mots)
 - proximité + ordre : pre/ (idem et respect de l'ordre des mots)
- opérateur de répétition : atleast (ex : *RESUME=atleast 5 sida*, recherchera toutes les phrases du résumé qui contiennent au moins cinq occurrences du terme *sida*).

Les parenthèses permettent d'ordonner les opérateurs booléens.

2.4. WWW-WAIS : SFgate

HTTP étant actuellement majoritaire sur Internet, des passerelles ont été créées pour interroger des bases WAIS.

SFgate en est un exemple, particulièrement adapté à l'interrogation de bases FreeWAIS-sf. Le CISM dispose de la version 4.0 (la dernière mouture est la 5.0).

Il s'agit d'un script CGI (voir page 77) écrit en Perl, qui permet d'interroger des bases WAIS au travers d'un formulaire d'interrogation écrit en HTML, donc disponible sur n'importe quel serveur HTTP. Il se comporte comme un logiciel client WAIS et suit le schéma suivant :

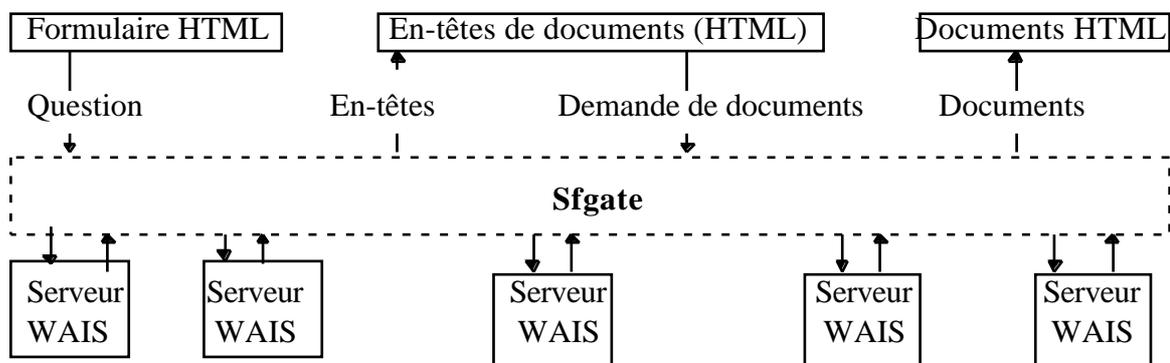


Figure 3 : fonctionnement de SFgate

2. 4. 1. Installation

Pour utiliser SFgate, il faut disposer au minimum de :

- un serveur HTTP qui lui fait appel
- un interpréteur Perl (voir page 78)
- le programme SFgate
- Les logiciels FreeWAIS-sf pour construire ses bases

Un module WAIS en perl est fortement conseillé pour accéder aux bibliothèques WAIS (*Waisperl* ou *Wais.pm*).

Il reste alors à décompacter les sources, installer les programmes sur le serveur et configurer l'ensemble (le tout laissé au bons soins du CISM, qui disposait depuis longtemps de cette passerelle).

2. 4. 2. La création de formulaires

L'interrogation des bases se fait par SFgate au moyen de formulaires. Il s'agit d'un code HTML réservé qui permet de faire appel à un programme, un script CGI (*Common Gateway Interface*), en l'occurrence la passerelle WWW-WAIS (mais les formulaires peuvent faire appel à n'importe quel autre programme).

Il faut indiquer :

- Le début et la fin du formulaire à l'intérieur du fichier HTML : `<FORM...>`
`</FORM>`
- La méthode d'envoi des données au serveur : *GET* ou *POST* (voir page 62)
- L'adresse IP du programme SFgate
- La base à interroger et ses coordonnées.

Il est ensuite possible de préciser :

- Les champs WAIS interrogeables : un seul champ (index global) ou plusieurs champs
- Les valeurs possibles dans un champ (valeur par défaut ou choix dans une liste)
- L'opérateur entre chaque champ, entre groupes de champs ou à l'intérieur des mots d'un champ (par défaut / choix dans une liste ou par cases à cocher)

- La façon dont seront affichés les résultats.

2. 4. 3. Syntaxe d'interrogation

La grammaire d'interrogation reste la même que celle de FreeWAIS-sf (voir plus haut). On peut rentrer les questions dans les zones du formulaire (une zone = un champ).

2. 4. 4. Visualisation des résultats

Lors de la requête, il est possible de choisir le nombre maximum de références en réponse et la manière dont elles seront affichés (lignes d'en-tête, ou directement le texte intégral). Après saisie de la question, on l'envoie en cliquant sur un bouton codé *submit* (soumettre la requête), et le tout est envoyé comme variable au script CGI qui interrogera la base.

Une fois la réponse obtenue, le principe est ensuite le même que tout client WAIS classique : on peut sélectionner les documents intéressants en cliquant sur leur ligne d'en-tête (voir exemple en annexe).

2. 5. Mise en place des données et du moteur sur le serveur

Le but du jeu était donc de récupérer les données de NosoBase, sous Basis plus, pour les transporter vers le serveur du CISM, où une base devait être créée sous FreeWAIS-sf.

2. 5. 1. Création de la base WAIS

La tâche principale qui nous incombait, puisque FreeWAIS-sf était déjà configuré et installé sur le serveur du CISM, était de décider d'une structure pour la base WAIS.

2. 5. 1. 1. Description des notices

En me basant sur un document trouvé sur Internet (*FREEWAIS-SF : Installation, Démarrage et Indexation* : voir Bookmark, rubrique *WAIS/FreeWAIS-sf/petit guide français*), ainsi que sur l'avis de l'administratrice WAIS du CISM, il m'a paru plus sûr de

remplacer les noms de champs Basis plus par leurs équivalents sur quatre lettres en majuscules : en effet, pour une meilleure indexation, il est préférable de délimiter correctement les champs. Les espacements et les noms de champs en sortie de Basis plus pouvaient être ainsi remplacés au profit de bornes plus précises.

La structure à indiquer dans le fichier de description des notices (*.fmt*) ne devrait donc pas provoquer d'erreur :

- un champ commence par quatre lettres en majuscules, suivis de deux points et d'un espace
- on trouve ensuite la valeur de ce champ
- il se termine par un saut de paragraphe suivi de quatre lettres en Majuscules, qui annoncent un autre champ. En effet, il y a théoriquement un champ par paragraphe, mais certains peuvent s'étaler sur plusieurs, comme les résumés ou les titres, ceci étant dû au formatage de sortie de Basis plus (insère des sauts de paragraphes à l'intérieur des champs longs pour une meilleure lisibilité)
- le séparateur des notices est un ensemble de soixante tirets, entre deux sauts de paragraphes
- les équivalents sur quatre lettres majuscules des noms de champs Basis plus sont les suivants (les contraintes de contenu sont les mêmes que celles expliquées en page 37) :

Intitulé sur la fiche manuelle d'indexation	Code Basis plus	Code de la base WAIS
n° d'indexation	#REF	REFE
Type de document	#TYPE	TYPE
Titre 1	#TITRE_FR	TIT1
Titre 2	#TITRE_OR	TIT2
Auteur(s)	#AUTEUR	AUTE
Adresse	#ADR_AUT	ADRE
Source	#SOURCE	SOUR
Editeur	#EDITEUR	EDIT
Date de publication	#DATE	DATE
ISSN	#ISSN	ISSN
ISBN	#ISBN	ISBN
Tome	#TOME	TOME

Volume	#VOLUME	VOLU
Numéro	#NUMERO	NUME
Page(s)	#PAGES	PAGE
Langue	#LANGUE	LANG
Pays	#PAYS	PAYS
Bibliographie	#BIBLIO	BIBL
Genre de document	#GENRE	GENR
Contenu	#MOTSCLES	MOTS
Contexte	#CONTEXTE	CONT
Résumé 1	#RESUME_FR	RES1
Résumé 2	#RESUME_OR	RES2
Commentaire (si nécessaire)	#COMMENT	COMM
-----	#INDEXEUR	INDE

Tableau 8 : champs WAIS de NosoBase

On trouvera en annexe un exemple de notice WAIS retournée par SFgate à l'interrogation du serveur.

2. 5. 1. 2. Le reformatage des données vers WAIS

Les données peuvent être récupérées à partir du logiciel client Basis plus et imprimées dans un fichier : le document résultant est au format Word, les caractères étant disposés de telle manière que les couples *nom du champ / valeur* soient tous bien alignés une fois le document imprimé. Il en résulte beaucoup d'espaces, de sauts de paragraphes insérés dans la notice en sortie.

Les notices ne pouvaient être présentées ainsi pour construire la base WAIS, car il est préférable de bien structurer les champs entre eux, pour une description plus sûre dans le fichier *.fmt* qui servira à l'indexation des données.

Il a donc fallu reformater de nouveau les données : celles en sortie de Basis plus devant se retrouver sous la forme WAIS précédemment citée.

2. 5. 1. 2. 1. Macros-commandes Word

Une première étape a été réalisée à l'aide de bonnes vieilles macros Word...

Un programme relativement simple permettait de remplacer :

- les noms de champs Basis plus par les noms WAIS
- les espaces afin que les notices soient correctement lisibles une fois retournées par SFgate

- le séparateur de notices
- les codes alphabétiques ou numériques des champs *Type de document* et *Genre du document* par leur valeurs explicites (ex : *O* remplacé par *Ouvrage*).

Le remplacement était effectué à l'aide de la fonction *EditionRemplacer*, identique à l'action provoquée par l'appel du menu *Edition, Remplacer* de Word.

Cette macro fonctionne, mais elle est beaucoup trop lente pour traiter l'ensemble actuel des notices : les 2 300 références représentent 4 Mo qui ne peuvent être traitées en temps raisonnable en une seule fois, et le fractionnement en petit groupe conduit à répéter les tâches, ce qui n'est pas admissible (le traitement de 300 notices prend 5 minutes sur un 486 DX2-66 !). Ceci est peut-être dû à l'utilisation de la fonction *EditionRemplacer*, qui lit et relit plusieurs fois le fichier (les accès au disque sont souvent dommageables à la vitesse d'exécution).

2. 5. 1. 2. 2. Clipper, langage de Microsoft Foxpro

Les macros ont cependant été utilisées pour un premier reformatage, mais à l'avenir cette solution n'est pas viable : il faut que la mise à jour de la base WAIS, donc un reformatage des données supplémentaires ou modifiées pour une réindexation, se fasse simplement et rapidement.

Je pensais créer un exécutable à l'aide d'un langage compilé comme le C++, ou Visual Basic. Louis AYZAC m'a proposé de faire le traitement en utilisant Microsoft Foxpro et Clipper, compilateur d'un langage similaire à Dbase (gestionnaire de fichiers). Après avoir décidé d'un autre traitement algorithmique des données, à priori plus rapide (pas d'accès multiple au disque), la structure était donc la suivante :

- le programme parcourt une fois le fichier, qu'il stocke dans une table (comme Dbase) : une entrée = une ligne du fichier
- un fichier en sortie est ouvert
- les douze premiers caractères de chaque ligne sont traités (ceux-ci sont réservés aux noms de champs en sortie de Basis plus) et, en fonction de leur valeur (*CASE OF*), une série de tests est effectuée. Puis, après écriture du

nom de champ WAIS dans le fichier de sortie, le contenu du champ est écrit en sortie comme tel, ou il est remplacé par une valeur plus explicite

- les espaces et sauts de paragraphes sont traités de la même façon.

Louis AYZAC, qui connaît très bien ce langage, a ajouté aussi quelques traitements qui facilitent le temps de chargement des données dans la table et donnent la durée de traitement.

Le résultat change radicalement : 4 minutes pour traiter le fichier de 4 Mo (environ 2 300 notices) sur le DX2 !

Le programme doit être encore "paufriné" (ajout d'un menu : formatage de toutes les notices, mise à jour seulement de quelques données, date de la dernière opération et nombre de notices traitées à cette occasion...). Il est une des clés du système.

2. 5. 1. 3. Indexation des fichiers

Une fois les données calquées sur notre modèle WAIS, elles doivent être transmises par FTP sur la machine du CISM. Un compte avait déjà été attribué pour Louis AYZAC afin d'abriter les pages du serveur de l'Université de Médecine Lyon Sud. Un répertoire *nosobase* a donc été créé pour y placer des notices : seule une petite partie (256 notices) a d'abord été stockée, car une mise à disposition intégrale de la base pour tout public (serveur Internet sans mot de passe) n'avait pas encore été décidé.

Il a fallu donner avec la description de la base (page 57) la composition de la ligne d'en-tête des documents, celle qui est retournée en résultat d'une interrogation : ce sera le contenu du champ *TITI* (*Titre français*) suivi, s'il reste de la place (la ligne est configuré pour seulement cent caractères) du champ *AUTE* (*Auteur(s)*).

Après plusieurs essais (mauvais format de texte [problèmes de caractères accentués], décalage peu esthétique des données...) et de nombreux contacts avec l'administratrice, l'échantillon a pu être ainsi indexé.

2. 5. 1. 4. Le fichier .src

Un fichier descriptif de la base a été aussi fourni, pour une éventuelle communication au *directory of servers* français, à l'INRIA, et lors d'une interrogation à l'aide de clients WAIS, ce qui ne devrait plus trop se faire actuellement.

2. 5. 2. L'interface d'interrogation SFgate

L'interrogation de bases WAIS se fait dorénavant par l'intermédiaire de formulaires HTML appelant des passerelles WWW*Erreur! Signet non défini.*-WAIS.

Le CISM abrite plusieurs bases WAIS, toutes interrogeables à l'aide de formulaires similaires : l'administratrice du CISM nous a donc proposé d'en créer un pour nous.

Nous lui avons fourni les champs qui nous semblaient intéressants à interroger (d'après l'expérience de Nathalie LEROY), et avons pu tester cette première base. Les premiers résultats s'avéraient satisfaisants.

Il ne me restait donc plus qu'à créer notre propre formulaire, plus "personnalisé", à abriter sur les pages d'un serveur.

2. 5. 2. 1. Le formulaire HTML/SFgate

Pour créer ce masque d'interrogation, je me suis appuyé sur les autres bases, ainsi que sur la documentation de SFgate 4.0, tous disponibles sur le serveur du CISM

Deux types de formulaires ont été montés : un premier très complet, pour interroger à l'aide de plusieurs champs, rentrer des questions en syntaxe WAIS..., puis un second, pour interroger simplement l'index global. On trouvera une copie papier en annexe.

2. 5. 2. 1. 1. Interrogation avancée

Le code HTML du formulaire suit ci-dessous. Les explications n'en font bien évidemment pas partie sur le serveur :

```
<FORM METHOD=GET ACTION="http://www.univ-lyon1.fr/cgi-bin/SFgate.new" target="hright">
```

FORM est la balise des formulaires. La *METHOD* peut-être *GET* ou *POST* (détermine le contenu de la variable d'environnement *REQUEST_METHOD*, qui stockera les informations passées par le formulaire). *ACTION* est l'adresse du script CGI (voir page 77) qui va traiter les données envoyées : il s'agit de SFgate. *TARGET* est une balise pour la "frame cible" (voir page 71).

```
<INPUT NAME="database"  
  TYPE="hidden"  
  VALUE="wais.univ-lyon1.fr/ROCAD-CCLIN-Nosobase">
```

NAME permet ici d'interroger la base de données spécifiée par *VALUE* (adresse de la base WAIS au CISM). Le *TYPE hidden* permet de conserver les données tapées par l'utilisateur après envoi, pour une réutilisation ou une comparaison par exemple.

<center>Les termes à l'intérieur des champs sont connectés par l'opérateur OU</center>

"Les termes à l'intérieur des champs sont connectés par l'opérateur OU"

<PRE>

la balise *<PRE>...</PRE>* permet d'afficher le texte tel qu'il est mis en forme sur le fichier ASCII original, du moins en ce qui concerne les espaces et les sauts de paragraphes (sinon il faut utiliser des codes HTML tels que *<p>* = saut de paragraphe, etc...). D'où des décalages sur le fichier source assez gênants pour la lecture, qui ont été modifiés pour l'explication.

Numéro de référence

<INPUT TYPE=TEXT NAME="refe" SIZE="30"><INPUT TYPE="hidden" >

"Numéro de référence" : *INPUT* est une entrée de formulaire. Le type est du *TEXT*, il concerne le champ WAIS *NAME (refe)*, et la taille *SIZE* de la zone de saisie est de 30. Enfin, le *TYPE hidden* permet de conserver cette donnée.

<SELECT NAME="type_tie">

La balise *<SELECT>..<OPTION>...* affiche une liste. La variable *SFgate* à prendre en compte ici (*type_tie*) est le connecteur (*_tie*) entre le champ précédent (*refe*), et le champ suivant (*type*) : *ET* ou *OU*

<OPTION>et

<OPTION>ou

</SELECT> Type de document

<SELECT TYPE=TEXT name="type"><INPUT TYPE="hidden" >

<OPTION SELECTED VALUE="">Tous

Même type d'explication qu'auparavant. *SELECTED VALUE* permet, bien évidemment, d'avoir une valeur sélectionnée par défaut. Il s'agit ici d'une liste des types de documents disponibles.

<OPTION VALUE="Ouvrage">Ouvrage

<OPTION VALUE="Péroratoire">Péroratoire

<OPTION VALUE="Méthode">Méthode

<OPTION VALUE="Rapport">Rapport

<OPTION VALUE="Congrès">Congrès

<OPTION VALUE="Législation">Législation

<OPTION VALUE="Autres">Divers

</SELECT>

<SELECT NAME="tit1_tie">

Opérateur entre le *type* est le *tit1 (titre français)*

<OPTION>et

<OPTION>ou

</SELECT> Titre français

<INPUT TYPE=TEXT name="tit1" SIZE="30"><INPUT TYPE="hidden" >

<INPUT TYPE="hidden" NAME="group_1" VALUE="aute_1,aute_2,aute_3">

<SELECT NAME="group_1_tie">

group_1_tie est l'opérateur entre le champ précédent (*tit1*) et le groupe indiqué ci-dessus. Ce groupe se définit dans les balises *INPUT* qui précèdent : *group_1* est le nom (*NAME*) du groupe, constitué des zones (*VALUE*) *aute_1*, *aute_2* et *aute_3*. Ces zones sont en fait plusieurs valeurs répétées du même champ WAIS : *aute* (*Auteur(s)*). Ce système permet ainsi de préciser plusieurs auteurs (une case par auteur dans le formulaire), et de grouper l'ensemble avec un opérateur booléen commun.

```
<OPTION>et
<OPTION>ou
</SELECT> Auteur(s)
<INPUT TYPE=TEXT name="aute_1" SIZE="12"><INPUT TYPE="hidden" > <SELECT
NAME="aute_2_tie">
```

1ère zone se rapportant au champ *aute* (*Auteur(s)*) : *aute_1*, qui possède aussi son propre opérateur avec la deuxième zone

```
<OPTION>et
<OPTION>ou
</SELECT> <INPUT TYPE=TEXT name="aute_2" SIZE="12"><INPUT TYPE="hidden" >
<SELECT NAME="aute_3_tie">
```

2ème zone pour le champ *aute* (*Auteur(s)*) : *aute_2*.

```
<OPTION>et
<OPTION>ou
</SELECT> <INPUT TYPE=TEXT name="aute_3" SIZE="12"><INPUT TYPE="hidden" >
```

3ème zone pour le champ *aute* (*Auteur(s)*) : *aute_3*.

```
<SELECT NAME="sour_tie">
```

Opérateur entre *group_1* (*Auteur(s)*) et le champ *sour* (*Source*)

```
<OPTION>et
<OPTION>ou
</SELECT> Source
<INPUT TYPE=TEXT name="sour" SIZE="30"><INPUT TYPE="hidden" >
```

Champ *sour* (*Source*)

```
<SELECT NAME="date_tie">
<OPTION>et
<OPTION>ou
</SELECT> Date de publication
<INPUT TYPE=TEXT name="date" SIZE="30"><INPUT TYPE="hidden" ><INPUT TYPE="hidden"
NAME="group_2" VALUE="mots_1,mots_2,mots_3">
```

Champ *date* et 2ème groupe (3 zones pour le champ *mots* [*Mot(s)-Clé(s)*])

```
<SELECT NAME="group_2_tie">
<OPTION>et
<OPTION>ou
</SELECT> Mot(s)-clé(s)
<INPUT TYPE=TEXT name="mots_1" SIZE="12"><INPUT TYPE="hidden" > <SELECT
NAME="mots_2_tie">
<OPTION>et
<OPTION>ou
</SELECT>
<INPUT TYPE=TEXT name="mots_2" SIZE="12"><INPUT TYPE="hidden" > <SELECT
NAME="mots_3_tie">
<OPTION>et
```

```
<OPTION>ou  
</SELECT>  
<INPUT TYPE=TEXT name="mots_3" SIZE="12"><INPUT TYPE="hidden" >  
<SELECT NAME="cont_tie">  
<OPTION>et  
<OPTION>ou  
</SELECT> Indexation libre  
<INPUT TYPE=TEXT name="cont" SIZE="30"><INPUT TYPE="hidden" >
```

Champ *cont* (Indexation libre)

```
<SELECT NAME="res1_tie">  
<OPTION>et  
<OPTION>ou  
</SELECT> R&eacute;sum&eacute;,fran&ccedil;ais  
<INPUT TYPE=TEXT name="res1" SIZE="30"><INPUT TYPE="hidden" >
```

Champ *res1* (Résumé français)

```
<SELECT NAME="text_tie">  
<OPTION>et  
<OPTION>ou  
</SELECT> Ensemble de la notice  
<INPUT TYPE=TEXT name="text" SIZE="30"><INPUT TYPE="hidden" >
```

Le nom (*NAME*) *text* est réservé pour interroger l'index global (indexation de l'ensemble des documents, tous champs compris).

```
<input TYPE="hidden" NAME="multiple" VALUE="1">
```

multiple permet de sélectionner plus d'un document à la fois parmi les lignes d'en-têtes retournées par le serveur (cases à cocher en face de chaque ligne).

```
<input NAME="directget" TYPE="checkbox" VALUE="1"> Cliquez ici pour acc&eacute;der  
directement aux documents
```

Case à cocher (*checkbox*) pour afficher directement les documents au lieu des lignes d'en-têtes.

```
<input NAME="maxhits" TYPE=TEXT value="40" size=2> Indiquez ici le nombre de documents  
recherch&eacute;s
```

Zone pour indiquer le nombre maximum (*maxhits*) de références à afficher en réponse à une question. La valeur par défaut est 40.

```
<input type="submit" value="Recherche"> <input type="reset" value="Effacer Recherche">
```

Il s'agit ici de deux "boutons", nommé *Recherche* et *Effacer Recherche*. Le premier envoie la question au serveur (*submit*), le second efface les données du formulaire (*reset*).

```
<INPUT TYPE="hidden" NAME="range" value="1">
```

Ceci permet d'afficher un lien *D'autres documents...* vers les lignes d'en-têtes suivantes, lorsque la réponse contient plus de références que le nombre demandé dans la zone *Indiquez ici le nombre de documents recherchés*.

```
</FORM></PRE>
```

fin des balises de formulaires (*FORM*) et de mise en forme (*PRE*).

...(suite du fichier)

2. 5. 2. 1. 2. Interrogation simple

Le principe est le même. Il n'y a qu'un champ, il s'agit de l'index global, codé ainsi dans le formulaire HTML :

```
<FORM METHOD=GET ACTION="http://www.univ-lyon1.fr/cgi-bin/SFgate.new" target="hright">
<INPUT NAME="database"
  TYPE="hidden"
  VALUE="wais.univ-lyon1.fr/ROCAD-CCLIN-Nosobase">
<input type="hidden" name="tieinternal" value="and">
```

tieinternal est le nom de la variable contenant l'opérateur booléen par défaut à l'intérieur du champ (*and*)

```
<center>
Entrez un ou plusieurs mots correspondant à votre recherche. Les termes sont connectés
par l'opérateur ET
</pre>
</pre>
<INPUT TYPE=TEXT name="text" SIZE="40"><INPUT TYPE="hidden" >
```

Il s'agit ici du seul champ interrogeable : l'index global (*name="text"*)

```
</pre>
</pre>
<input NAME="directget" TYPE="checkbox" VALUE="1"> Cliquez ici pour accéder
directement aux documents
</pre>
</pre>
<input NAME="maxhits" TYPE=TEXT value="40" size=2> Indiquez ici le nombre maximum de
documents recherchés
</pre>
</pre>
<input type="submit" value="Recherche"> <input type="reset" value="Effacer Recherche">
<INPUT TYPE="hidden" NAME="range" value="1"></FORM>
```

2. 5. 2. 2. Exemple d'interrogation avancée

Les "étapes-écran" de cet exemple sont disponibles en annexe.

Le formulaire est rempli pour poser la question suivante :

Mot(s)-Clé(s) : SIDA ET MATERNITE

Une fois le bouton *Recherche* enfoncé, la question est transmise à SFgate, sur le serveur du CISM. Celui-ci recherche dans la base construite avec FreeWAIS-sf, et retourne les lignes d'en-tête des documents correspondants à la question, classés par ordre décroissant de pertinence (principe de pondération WAIS).

Le nombre affiché correspond au nombre maximum demandé à l'interrogation (si la réponse en contient plus, un lien *D'autres documents...* apparaît en fin d'écran). Le principe d'affichage des documents peut varier d'un formulaire à l'autre, et selon les choix effectués par l'utilisateur.

Il suffit ensuite de cocher les documents qui nous semblent intéressants, de cliquer sur *rechercher les documents*, et le texte intégral est retourné par le serveur.

2. 5. 2. 3. Un produit perfectible

L'interrogation de la base WAIS fonctionne de cette façon, et est satisfaisante pour une première mise en place. Quelques restrictions sont cependant à faire :

- Les opérateurs mentionnés dans la documentation de FreeWAIS-sf ne fonctionnent pas tous : ET, OU, * et les parenthèses donnent les résultats attendus (bien que l'ordre des opérateurs ET/OU ne soit pas respecté s'il ne sont pas employés avec des parenthèses) ; mais les opérateurs de proximité (soundex, w/, pre/ : voir page 53), de répétition (atleast), les doubles côtes (expression littérale) et le NOT (sauf) ne fonctionnent pas. Le plus gênant est surtout l'absence du NOT, le reste permettant des interrogations plus poussées qui, de la part d'utilisateurs non spécialisés (non documentalistes), seront certainement rares.

Il semble que les utilisateurs des autres bases WAIS du CISM n'aient pas eu de tels besoins, puisqu'ils ne fonctionnent pas non plus chez eux

- Les caractères accentués (*ISO 8859 Latin-1*) ont longtemps été la "bête noire" des internautes des pays latins. Un *patch* WAIS (programme ajouté pour corriger des bugs ou ajouter des fonctionnalités) a pourtant été créé pour pallier cette difficulté, les bases pouvant donc contenir des accents.

Ce n'est pas le cas pour la notre, ce qui produit des réponses pas toujours en rapport avec la question. Ainsi, l'emploi des termes *seropositivite* et *séropositivité* ne donnera pas le même nombre de réponses. Cela est flagrant pour le champ *Mot(s)-Clé(s)*, qui reprend les mots du thesaurus des C.CLIN toujours en majuscules : en questionnant sur ce champ avec des accents (donc en minuscules), les documents obtenus ne contiennent pas toujours le mot-clé demandé !

- Quelques erreurs d'indexation ont été relevées. En questionnant seulement le champ *Type de document* avec la valeur *Thèse ou mémoire*, on obtient sept lignes d'en-tête : il y a bien les quatre mémoires de la base, mais les trois

autres contiennent les termes *Thèse* ou *Mémoire* dans un autre champ, non demandé !

- Enfin, une fois la requête envoyée à SFgate, nous n'avons plus la main sur les documents retournés : l'en-tête et le pied de page sont des liens vers le manuel de SFgate, il n'est pas possible de récupérer une information sur un document ni d'insérer un lien HTML (*Commander cet article*, par exemple), etc... Ceci est possible, bien sûr, mais il faudrait se plonger dans des modifications de configuration de la passerelle, ce qui n'a pas été envisagé, car :

il n'y a qu'un SFgate pour toutes les bases abritées sur le serveur du CISM. Il faudrait donc nous créer un programme "personnalisé", ce qui n'a pas été demandé pour l'instant

FreeWAIS-sf est un programme gratuit, certes, mais il faut respecter leurs auteurs. L'administratrice du CISM ne préfère pas supprimer le lien vers le manuel de SFgate, ou faire trop de modifications. Cela nous obligerait donc à configurer le programme nous-même, et la modification de SFgate n'est pas notre priorité : il faut déjà que l'interrogation fonctionne.

La base n'a donc pas donné l'exact résultat que nous attendions. Cela vient peut-être de la version de FreeWAIS-sf ou de SFgate, en tout cas les manuels ont été suivis à la lettre.

De notre côté, il semble que nous ne puissions rien faire pour améliorer ces défauts. Nous attendons le retour de vacances de l'administratrice du CISM pour tenter de supprimer ces problèmes.

Il reste cependant que l'interrogation fonctionne quasi-correctement, c'est-à-dire que les réponses sont la plupart du temps en accord avec les questions posées, lorsque l'on est bien averti des outils à ne pas utiliser (accents, opérateurs). Le système a été testé, par Nathalie LEROY notamment qui connaît bien sa base, et ce premier essai s'avère concluant, bien que perfectible.

N'oublions pas aussi que WAIS est un bon outil pour interroger des bases documentaires sur Internet, mais qu'il n'est pas, à l'heure actuelle, au niveau des produits documentaires spécialisés, comme Basis plus notamment. Au grand regret des documentalistes...

2. 5. 3. Le serveur WWW NosoBase

Bien que le CISM puisse abriter notre formulaire, il nous paraissait important d'avoir nos propres pages, pour y intégrer un guide, des informations sur les C.CLIN, etc...

Mise en place le 28 Août 1996, il est disponible à l'adresse suivante (respectez les majuscules/minuscules) :

<http://www.univ-lyon1.fr/LyonSud/nosobase/departf.htm>

Nous ne reprendrons donc pas en détail le contenu, visible sur Internet (copies du contenu en annexe, la forme n'étant pas imprimable dans les moindres détails).

2. 5. 3. 1. Développement HTML

2. 5. 3. 1. 1. Editeurs utilisés

Environ un tiers du travail a été réalisée sous l'éditeur HTML Mozilla fourni avec Netscape Gold 2.01. Cet outil WYSIWYG ("What You See Is What You Get" : on obtient à l'écran ce qui se verra sur le serveur) ne manipule pas les tableaux (la version 3.0 Beta corrige ce manque), les frames, le clignotement, ... mais offre des facilités pour la mise en forme du texte (gras, italique, taille, couleur), les listes (ordonnées ou non), le décalage des caractères, la création de liens, d'ancres, l'insertion d'images... Il est très pratique, mais ne suffit pas seul.

Une grande partie du développement a donc été réalisé "à la main", avec un éditeur de texte (Bloc notes ou Word) en reprenant le code HTML. A ce propos, la relecture du code généré par Mozilla est souvent un calvaire (insertion de balises qui annulent d'autres balises, etc...), et il est nécessaire de faire le nettoyage .

Les débuts en HTML sont assez contraignants, mais l'on s'y fait vite, et les résultats obtenus "de visu" donnent envie de continuer.

Je me suis beaucoup aidé de deux ouvrages présents dans la bibliographie, qui contenaient toutes les balises nécessaires, ainsi qu'en visualisant les codes sources de serveurs Internet.

Le tout a été développé en local (disque dur), puis testé sur le serveur du CISM.

Enfin, j'ai relativement vite compris les limites d'HTML en ce qui concerne l'interactivité, la mise en page (d'où l'apparition de nouveaux langages comme Java, VRML...), et les contraintes que représentent l'absence de contrôle de sa mise en forme. Il revient en effet à chaque utilisateur Internet de décider de la police présent sur son logiciel client, de la taille, des couleurs, ... sans compter les modifications apportées selon qu'on utilise un écran 14" ou 17". Il faut donc distribuer ses informations de la manière la plus lisible possible, pour une configuration de base. Ainsi, une version sans frame (voir ci-après) et le codage HTML des accents ont été implémentés.

2. 5. 3. 1. 2. Particularité HTML : les frames

Il est possible de scinder l'écran du logiciel client en plusieurs fenêtres indépendantes quant au contenu, appelées "frames". Dans chaque frame, on trouve un document HTML classique. Tout utilisateur peut bénéficier de ce multi-fenêtrage à condition de disposer de Netscape dans sa version 2 minimum, ou de Internet Explorer (versions récentes).

Pour notre serveur, cela nous a ainsi permis de disposer d'une barre de menus : celle-ci reste constamment à l'écran dans une frame, pendant qu'une autre frame reçoit les informations demandé par l'utilisateur qui a cliqué sur une des icônes du menu. La possibilité d'avoir toujours un contrôle sur son déplacement dans nos pages est ainsi offerte. Certains serveurs ont récemment supprimés ces frames, car elles rendaient leurs pages trop complexes (site de Netscape notamment). Pour nous, cette possibilité a semblé plus un avantage qu'un inconvénient, notamment pour garder la main lorsqu'une requête était envoyée (pour consulter le guide, par exemple). Une version sans frame est aussi disponible sur notre serveur.

Le déplacement dans une frame (retour, en avant) se fait différemment de l'habitude des documents mono-fenêtrés : il faut utiliser le bouton de sa souris (droit sur

PC) et choisir *Back in Frame*, ou *Forward in Frame* (au lieu du *Back* habituel sur la barre d'outils).

2. 5. 3. 1. 2. 1. Syntaxe

Un petit didacticiel HTML des frames a été trouvé sur le site de Netscape (voir Bookmark). Voici les points importants :

- La structure de base d'un document avec frame est la même qu'un document HTML classique, la balise `<BODY>` étant remplacée par `<FRAMESET>` :

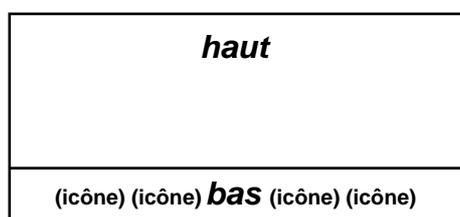
```
<HTML>  
  
<HEAD>  
</HEAD>  
  
<FRAMESET>  
</FRAMESET>  
  
</HTML>
```
- Dans la balise `<FRAMESET>`, on précisera si les deux frames sont horizontales (*ROWS*) ou verticales (*COLS*), ainsi que leur taille (*VALUE* en pixel ou pourcentage)
- A l'intérieur de la balise `<FRAMESET>` se trouvent deux autres balises `<FRAME>`, qui indiquent ce que vont contenir chacune des deux frames. On précisera ainsi le document HTML (*SRC*), un nom pour la frame (*NAME*) afin d'indiquer plus tard que c'est dans celle-ci que devront s'afficher les liens (*TARGET*), les marges (*MARGINHEIGHT* et *MARGINWIDTH*), si une barre de défilement s'affichera (*SCROLLING*) et s'il est possible de redimensionner la fenêtre (*NORESIZE*)
- Enfin, après fermeture par `</FRAMESET>`, une suite de balises `<NOFRAME>...<BODY>...</BODY></NOFRAME>` permet de stocker du texte pour les visualiseurs clients anciens, qui ne peuvent afficher les frames (Netscape 1 par exemple)
- A l'intérieur d'une frame, donc du document HTML qui la compose, il est possible de définir une "frame cible" pour recevoir les documents appelés par des liens (pour permettre par exemple qu'un clic dans la barre de menus dans la frame du bas, affiche le résultat dans la frame du haut, ce afin de

conserver constamment cette barre de menus à l'écran) : ceci se fait à l'aide de l'attribut *TARGET*, à l'intérieur des liens (*A HREF="url" TARGET="nom de la frame affecté dans la balise <FRAME>...*), d'une carte, d'un formulaire ou pour tout le document (*BASE TARGET*).

En effet, une frame contient un document HTML classique, qui peut être lui aussi composé de frames (d'où des imbrications de fenêtres dans des fenêtres dans des... Mais le nombre total de frames par fichier HTML est de 2).

2. 5. 3. 1. 2. 2. Exemple

La page de garde de notre serveur est composé de deux frames horizontales.



Graphique 6 : les frames dans *departf.htm*

Le document appelé pour afficher ce découpage s'appelle *departf.htm* ("depart avec frame" : le point d'entrée de notre serveur). Les lignes suivantes donnent un aperçu du code source concernant les frames.

```
<FRAMESET ROWS="*,63">
```

Le document comporte des frames disposées en ligne (*ROWS* : l'une en dessous de l'autre), la seconde mesurant 63 pixels de haut et la première prenant le reste de la place.

```
<FRAME NAME="haut" SRC="bienvenuf.htm" SCROLLING = "yes" NORESIZE>
```

La première frame (du dessus) porte le nom de *haut* (*NAME*), elle est constituée du document HTML *bienvenuf.htm* (*SRC*), d'une barre de défilement (*SCROLLING*), et sa taille n'est pas modifiable (*NORESIZE*)

```
<FRAME NAME="bas" SRC="menubarf.htm" SCROLLING = "no" NORESIZE>
```

La seconde frame (du dessous) porte le nom de *bas*, elle est constituée du document HTML *menubarf.htm*, d'une barre de défilement, et sa taille n'est pas modifiable.

```
</FRAMESET>
```

```
<NOFRAME>
```

Contient le texte qui sera affiché pour les anciens "browsers".

```
<BODY...>
```

```
...
```

Ce serveur utilise des frames (pages avec plusieurs fenêtres) et nécessite au moins Netscape 2.0 pour être visualisé

```
</BODY>
```

```
</NOFRAME>
```

```
</HTML>
```

La frame du dessous, nommée *bas* contient le document *menubarf.htm* (barre d'icônes), qui comprend la balise `<BASE TARGET="haut">`. Ainsi, lorsqu'un clic sur un icône sera effectué, le document appelé apparaîtra dans la frame du haut du document *departf.htm*.

Le fonctionnement des frames est donc très simple.

2. 5. 3. 2. "Philosophie" du serveur

Sur Internet, malgré les contraintes du langage HTML, on trouve des serveurs très bien réussis esthétiquement, avec des photos, des cartes pour naviguer entre les pages, des icônes animés, etc... La contrepartie est parfois un temps de chargement assez long lorsque le réseau est saturé, ou en utilisant un poste lent.

Notre but est avant tout de mettre à disposition une base de données, pas de faire un serveur à la pointe du langage HTML.

J'ai donc souhaité proposer une présentation claire, sobre, avec un temps de chargement raisonnable. Les premières impressions qui nous ont été exprimées vont dans ce sens.

Quelques images (lignes séparatrices, deux fonds d'écran et icônes) ont été intégrées, ce qui permet de sortir du gris un peu triste des pages HTML basiques mais ne ralentit pas trop la vitesse de rapatriement des données. A part les frames, qui sont une aide à la navigation plus qu'un gadget, aucune autre sophistication n'a été tentée.

De plus, deux sortes d'interrogation sont disponibles, pour un public novice (la plupart des internautes certainement) et un autre plus averti.

Enfin, plusieurs documents sont présents pour renseigner l'utilisateur (guide, informations sur la base...).

2. 5. 3. 3. Contenu

Dans la version avec frame, une barre de menus est proposée en bas de l'écran, tandis que le haut contient le contenu des différentes rubriques présentées en bas.

Dans le modèle sans frame, un alignement d'icône à la fin de chaque document permet d'atteindre les autres rubriques.

Chaque document, dans les deux versions, contient des liens hypertextes internes (ancres) ou externes (renvoient vers d'autres documents).

Voici les différentes rubriques proposées sur notre serveur (copies en annexe) :

- *Interrogation simple*
- *Interrogation avancée*
- *Guide* : il s'agit d'un document de référence pour utiliser correctement les formulaires d'interrogation. Il contient les points suivants :

Utilisation : explication sommaire pour interroger la base

Champs interrogeables : champs présents dans les formulaires HTML (intitulé, contenu)

Syntaxe d'interrogation : interrogation des champs (exemples), avertissement sur le problème des accents

Opérateurs : présentation des opérateurs disponibles

Format des notices : comment lire une notice de document telle qu'elle sera retournée (nom des champs WAIS)

Consulter les références : donne les coordonnées pour consulter les références trouvées. Il a été décidé pour l'instant que les utilisateurs souhaitant obtenir les documents référencés dans la base devront s'adresser au C.CLIN Sud-Est (les coordonnées sont fournis, ainsi qu'un lien HTML qui permet d'envoyer un courrier électronique) : si Nathalie LEROY ne possède pas les documents, elle pourra les aiguiller vers d'autres C.CLIN. De plus, la situation vis-à-vis des droits d'auteurs n'ayant pas été encore arrêtée, le

terme *Consulter les références* a été utilisé sur le serveur, plutôt que *Commander* ou *Obtenir*

- Présentation : informations sur NosoBase (contenu, définition des infections nosocomiales, les C.CLIN [statut, coordonnées], la base WAIS [principes, versions utilisées, serveur du CISM])
- Bienvenue : il s'agit de la page de garde (*homepage*) du serveur. Elle est composée :

d'avertissements sur la nécessité de disposer de Netscape 2 pour utiliser les frames et sur l'utilisation des accents (car une supposition est que peu de personnes prendront vraiment le temps de lire le guide)

d'un lien vers la version sans frame

d'une présentation de nos rubriques

de liens pour consulter les références (vers le guide)

de contacts par courrier électronique vers Louis AYZAC, Nathalie LEROY, le CISM et moi-même

enfin, de la date de dernière mise à jour.

2. 5. 3. 4. Lancement du serveur

Un rapport de toute la chaîne WAIS-WWW est en cours de réalisation pour laisser un "mode d'emploi" après mon départ.

2. 5. 3. 4. 1. Ouverture au public

Le serveur devait être au départ pour les cinq C.CLIN de France. La décision de l'ouvrir sur l'Internet, donc pour tout le monde, devait se poser plus tard.

Une restriction des utilisateurs suppose un Intranet, c'est-à-dire un réseau local fermé à l'Internet par des *Firewalls*, ou la saisie d'un mot de passe pour se connecter. D'où une charge de travail supplémentaire pour sécuriser un tel système.

Un des obstacles à l'ouverture "tout public" était un projet parallèle de CD-ROM sur les infections nosocomiales, avec la société Digital Média : il contiendrait NosoBase, ainsi que d'autres documents. L'idée de départ était d'articuler Internet et le CD-ROM, le CD comportant NosoBase en l'état actuel et le serveur permettant de télécharger les mises à jour (répartition *base fixe / information plus fraîche*, que l'on retrouve de plus en plus, comme dans le CD *Microsoft Encarta* par exemple). Quelques réticences étaient émises, notamment celle de ne pas pouvoir faire face financièrement à l'activité supplémentaire qu'engendrerait le serveur Internet. Il fallait donc vendre le CD, et le serveur pouvait lui faire de l'ombre (pourquoi les gens achèteraient le disque alors que les informations sont disponibles gratuitement sur Internet ?).

Finalement, il a été décidé que le CD ne reprendrait pas NosoBase en tant que base de données, mais présenterait un contenu bien différent, non concurrent (...donc vendable !). Après quelques âpres discussions, la mise à disposition de NosoBase sur Internet a donc été acceptée.

2. 5. 3. 4. 2. Une première phase de test avant la promotion

Les fichiers ont donc été transporté par FTP sur la machine du CISM et, après quelques rectifications (les images n'apparaissaient pas, car j'avais oublié dans les noms de fichiers correspondants qu'Unix faisait la différence majuscules/minuscules !), tout semblait fonctionner.

L'ouverture de nos pages n'a d'abord été annoncée que dans le pavillon 1 M, aux quatre autres C.CLIN et, cela va de soi, sur la liste des élèves du DESS !

En effet, l'interrogation n'est pas sans défaut, nous l'avons déjà dit, et des erreurs, manques de clartés, dysfonctionnements divers peuvent nous avoir échappés.

Nous attendons donc la fin du mois de Septembre (retour de l'administratrice du CISM) pour "annoncer officiellement l'apparition du nouveau-né", sur les listes de discussions spécialisés, les moteurs de recherche (ajout d'une balise HTML), forums, par courriers aux correspondants du C.CLIN, dans des revues comme *HygièneS*, etc...

2. 5. 3. 5. Une charge de travail supplémentaire

Nathalie LEROY et le personnel du C.CLIN Sud-Est sont bien conscients du travail supplémentaire que va apporter ce serveur. Par ailleurs (est-ce un hasard ?) deux personnes pour NosoBase viennent d'être embauchées. Mais cela permettra sans aucun doute d'accroître leur public, donc de mieux faire connaître la base et les C.CLIN.

2. 6. Saisir les données par Internet : l'interface CGI

Une dernière étape prévue au départ était la possibilité offerte aux C.CLIN de saisir leurs notices directement par Internet, en se connectant sur notre serveur.

A la fin du mois de Juillet, une étude a été menée dans cette optique, puis d'autres priorités sont apparues. Cette facilité n'est donc pas disponible actuellement, mais voici quelques points de départ qui permettront de la mettre en place.

2. 6. 1. Principes CGI

CGI (*Common Gateway Interface*) est un standard pour interfacier des documents Internet avec des programmes tournant sur les serveurs (appelés scripts CGI). Il permet par exemple d'interroger une base de données : le formulaire HTML est rempli par le client, envoyé au serveur par une passerelle CGI qui déclenche un programme (script) pour interroger la base, puis renvoie le résultat au client (SFgate est ainsi un script CGI).

Il s'agit donc d'une passerelle (*gateway*), utilisable pour n'importe quel programme (envoi automatisé de mail, stockage de données, interrogation de bases, inscription à une manifestation, enquête de satisfaction, ...).

Les scripts CGI sont toujours stockés dans un même répertoire (souvent *cgi-bin*). Le serveur "sait" ainsi qu'au lieu de renvoyer les documents vers le browser, il faut exécuter le programme.

L'utilisation de CGI suppose un travail de sécurisation des données sur le serveur (droits d'accès aux scripts et autres répertoires), puisque tout client peut faire tourner un programme sur la machine hôte, ce qui peut s'avérer dangereux.

Les programmes peuvent être écrits dans toutes sortes de langages supportés par le serveur. Il s'agit en général de Perl, Shell Unix ou C/C++, pour lesquels on peut trouver de nombreuses ressources (bibliothèques de fonctions) sur Internet.

2. 6. 2. le langage Perl

Le serveur du CISM possède entre autres Perl dans sa version 4.0, sur lequel je me suis donc penché.

Plusieurs manuels sur Perl sont disponibles sur Internet .

Il s'agit d'un langage assez "facile" (plus que le C), interprété (comme beaucoup de langages utilisés avec CGI) et qui propose des fonctions très puissantes pour manipuler du texte.

Un programme Perl peut contenir un appel à l'interpréteur (`#!/usr/local/bin/perl`), des variables "scalaires" (chaîne de caractères et numériques indifféremment : `$var=9`), des fonctions d'entrée/sortie (`print`), des opérateurs, tableaux (`@`), manipulations de fichiers (`open`), structures de contrôle (`foreach, for, while, until, if`), etc... et les fameuses expressions réservées (*regular expressions*), comme on en trouve en Unix.

Il s'agit de codes, composés souvent d'un caractère, utilisés pour manipuler les chaînes de caractères (identification, substitution, translation, découpage dans un tableau), d'où la puissance du langage. Souvent assemblées, elles produisent des lignes parfois illisibles, en tout cas pour un non averti. Exemple : `\|s*0\.0*` représente une division (`/`) par zéro avec des espaces possibles (`s*`), des décimales (`\.`) et quelques zéros (`0*`) qui peuvent suivre ! Avec un peu d'entraînement, on arrive à déchiffrer ces messages venus de l'au-delà...

Quelques programmes ont été écrits, à partir d'exercices présents dans des documents pédagogiques. Mais le script de saisie n'a pas été encore amorcé, faute de temps.

2. 6. 3. Appel du script : formulaire et droits d'accès

Une fois le programme créé et testé, il faut l'appeler à partir du logiciel client.

Cela se fait par un formulaire HTML, qui enverra les couples variables/valeurs de l'utilisateur au serveur. Nous ne revenons pas sur ce type de document (voir page 61), l'adresse référencée dans la balise `<FORM...>` doit être celle du script sur le serveur.

En ce qui nous concerne, le formulaire devra servir aux C.CLIN (clients) pour saisir leurs notices, qui nous seront envoyées (serveur) afin d'être vérifiées par le programme (script Perl) puis stockées sur la machine du CISM avant d'être transmises aux HCL (Basis plus) et à la base WAIS. Il est donc bien évident que seuls les C.CLIN pourront utiliser ce service. Plusieurs restrictions sont possibles :

- Créer un formulaire HTML qui contient toutes les informations à fournir au script CGI, et le distribuer uniquement aux C.CLIN (qui le copieraient donc en local sur leur disque dur). Ainsi, personne d'autre ne disposera du formulaire et ne saura ce qu'attend le programme. Même s'il est disponible aux yeux de tous, il ne pourra donc pas s'exécuter correctement et venir stocker ses données sur la machine du CISM
- Poser des droits d'accès sur la machine du CISM. Un formulaire de test a été placé dans un répertoire *saisie* situé juste en dessous du point d'accès à notre serveur. Un fichier *.htaccess* y a été créé, qui est lu avant *access.conf* du serveur : il contient entre autres une liste des utilisateurs autorisés dans le répertoire *saisie* (possibilité de créer un groupe), les droits d'écriture, le fichier qui contient les paires *nom utilisateur/mot de passe...* En cas d'accès au formulaire, une fenêtre est renvoyée automatiquement au client par le serveur, lui demandant de rentrer son identité (*login/password*).

Pour l'instant, seuls des tests ont donc été effectués, en ce qui concerne l'accès et le programme sous-jacent. La véritable création et mise en place n'a pu être réalisée, faute de temps.

3. Protection de "NosoBase" à l'INPI

Lors d'une réunion bi-mensuelle du C.CLIN fin juillet, Dominique Monnet m'a demandé de me renseigner quant aux démarches à suivre pour protéger leurs noms de produits : il était à l'époque question de décliner la dénomination *Noso* en trois appellations, *NosoBase*, *NosoRom* (le CD en projet avec Digital Média) et *NosoNet* (serveur Internet). Peut-être aussi *NosoNews*, pour la revue bibliographique.

Je me suis donc rendu à l'INPI (Institut National de la Propriété Industrielle) pour me renseigner.

3.1. Protéger ses noms : où ?

La protection se fait à l'INPI (Institut Nationale de la Propriété intellectuelle). Il existe 10 centres régionaux, dont :

INPI LYON

43, rue Raulin, 69 364 LYON Cedex 07

Tél : 78 72 59 42 - Télécopie : 78 61 77 21

Documentation importante sur place et renseignements (ingénieurs INPI).

3.2. Comment ?

Protéger son nom, c'est le faire au titre d'une marque. Ceci permet de se défendre contre toute similitude intellectuelle, phonétique ou écrite (ainsi, *NosoRom*, nom pressenti pour le CD, ne pourra être déposé, car il existe déjà *NosocoRom*, même type de produit).

La protection est valable dix ans à compter du dépôt, renouvelable indéfiniment (respecter les délais sous peine d'amende), en France. Elle est possible dans les états de l'Union Européenne (accord de Madrid) moyennant finances. Pour les autres pays (USA et Canada, notamment pressentis comme "concurrents" potentiels), il faut effectuer un dépôt sur le territoire même.

La protection se fait pour trois classes de produits (chaussures de sports, balais, ...) ou de services (base de données, logiciel...). Il faut payer au-delà. Les noms dérivés de *NosoBase* concernent les classes 9 (CD), 16 (Papier), 38 (Communication par terminaux d'ordinateurs), 41 (édition de livres, revues ; colloques ; éducation, formation...) et 42 (Bases de données).

La marque, cela peut-être l'association d'un nom et d'un logo : les deux sont alors protégés, ensemble et individuellement.

Il faut déposer chacun des noms (donc trois dépôts pour trois noms) pour les protéger individuellement : un dépôt commun ne les protégerait pas correctement. Cependant, seule la marque *Nosobase* peut être déposée, donc protégée (pour les cinq classes), et les autres supports utiliseraient alors ce nom. Attention : le radical *Noso* n'est pas assez distinctif à lui seul pour être déposé, puisqu'il est utilisé en médecine. Il faut donc bien protéger un nom distinctif, comme *NosoBase*.

3. 2. 1. Recherche d'antériorité

Il est nécessaire d'effectuer préalablement une recherche d'antériorité du nom, pour vérifier qu'il n'a pas été déjà déposé.

Voici le résultat des premières démarches effectuées en Juillet 1996.

- Fichier Marque de l'INPI

Recherche sur place (Base *MARLYS* sur Minitel : marques françaises depuis 1986). Résultats :

Noso(co)reseau(x), Noso(co)Network, NosoNew(s), Nosocomial, Nosobase, Nosonet, Nosorom, Nosocomi, Nosocomiaux, Nosocomix, NosoCD, Nosocome, Nosocum : aucune marque correspondante
Nosoco-Rom : (classes protégées : 35, 38, 41) CD-ROM ;
"concurrent" direct
NOSOCO : (classes protégées : 35, 36, 39, 40, 42) Société de service (finance, ingénierie...) ?
NOSOCOM : quel produit ? (classes protégées : 9, 42)
LES NOSOCOMIALES : revue ? (classes protégées : 16, 41)

NOSOCOMIA : classe ne nous concernant pas

- Bases de données en ligne :

ICIMARQUES : Minitel 36 29 36 30 (marques)

NosoBase, NosoNet, NosoNews, NosoRom, NosoCD : pas de dépôt similaire

Recherche sur le radical *NOSO* : 39 réponses dont 2 intéressantes, *Nosocom* et *Nosoco*

36 17 EURIDILE : dénominations sociales

Recherche sur *Nosobase* : aucune société à ce nom.

Il existe aussi *FMARK* et *TMINT* : sur abonnement.

N'ayant rien trouvé à ce niveau, il est indispensable de demander une recherche plus poussée à l'INPI, moyennant :

- 250 F / classe ; 125 F / classe supplémentaire
- 250 F / raison sociale ; 125 F / raison sociale supplémentaire.

Il faut en effet effectuer une recherche sur les noms de raisons sociales (Registre National du Commerce et des Sociétés), qui peuvent être un obstacle au dépôt d'une marque similaire. Elle se fait pour un GAS (Groupe d'Activités Similaires) : en ce qui nous concerne il s'agit des *services médicaux* (37) et de *l'enseignement* (38)

3. 2. 2. Dépôt

Si le nom s'avère alors disponible, on peut s'engager dans la procédure de dépôt :

- Dépôt à l'INPI ou au greffe du tribunal de commerce dont relève le domicile
- Paiement des redevances :

1er dépôt jusqu'à 3 classes	1 250 F
Par classe de produits ou de services (au-delà de 3)	250 F

Tableau 9 : tarif général de dépôt à l'INPI

Le renouvellement du dépôt après 10 ans coûte actuellement 1 400 F/3 classes + 250 F/classe supplémentaire

- Demande d'enregistrement en cinq exemplaires (imprimés INPI), avec :
 - identification du demandeur
 - représentation de la marque
 - énumération des produits ou services auxquels elle est destinée à s'appliquer (très important : détermine l'étendue de la protection)
 - indication des classes dont relèvent ces produits ou services
 - brève description de la marque et de ses éventuelles couleurs.

Cette demande doit être effectuée avec le plus grand soin. On pourra s'appuyer sur d'anciens dépôts comme modèles.

Elle peut être effectuée par un cabinet de conseil en Propriété Industrielle (compter quelques milliers de Francs).

3.3. Quand ?

- Date de dépôt = point de départ de la protection - Jour J
- Publication au BOPI (Bulletin Officiel de la Propriété intellectuelle) : J + six semaines
- Délai d'opposition : d'autres personnes peuvent faire opposition à ce dépôt dans les deux mois qui suivent la publication au BOPI (entreprises concurrentes, nom déjà déposé...)
- Examen par l'INPI
- Enregistrement (si recevable) : J + 4 Mois

3.4. Pourquoi ?

La marque (le nom) est un point de repère essentiel pour le "client". Elle engage la réputation de l'entreprise, c'est le symbole de son activité. Il s'agit d'un outil de promotion.

Une marque déposée doit être utilisée (sinon la protection peut être remise en question). De plus, il devient nécessaire de surveiller (Veille) si le nom n'est pas usurpé, imité...

Cependant, le droit des marques est un droit incertain : il est possible qu'un nom soit déjà utilisé, malgré des recherches infructueuses (l'INPI n'engage pas sa responsabilité dans les résultats de sa recherche d'antériorité). Il est essentiel de bien choisir son nom, sa politique marque en fonction de sa politique "commerciale". C'est la direction qui prendra la responsabilité du choix de la marque.

3. 5. En résumé...

1. Choisir un ou plusieurs noms
2. Demander une recherche d'antériorité à l'INPI (marque et registre du commerce)
3. Si rien de similaire n'a été trouvé, protéger au choix un nom (pour tous les produits) ou plusieurs (un dépôt par nom)
4. Protection pour cinq classes : 9, 16, 38, 41, 42
5. Remplir la demande de dépôt en cinq exemplaires avec une grande attention

Coûts pour un nom :

- Recherche d'antériorité :

Cette recherche peut se faire sur un nom seulement, et non pour tous, dans le cas où les différents noms se ressemblent (*Nosobase*, *NosoNet...*).

Une première recherche s'avérerait peut-être utile pour les autres.

Recherche d'antériorité à l'INPI	Marque : 250 F/classe 125 F/classe supplémentaire 5 classes (9, 16, 38, 41, 42)	750 F
	Registre du Commerce : 250 F par groupe d'activités 125 F par groupe supplémentaire 2 groupes (37, 38)	375 F
Total		1 125 F

Tableau 10 : tarifs pour une recherche d'antériorité de *NosoBase* à l'INPI

- Dépôt

1er dépôt jusqu'à trois classes	jusqu'à 3 classes	1 250 F
	250 F par classe supplémentaire 2 classes supplémentaires	500 F
Total		1 750 F

Tableau 11 : tarifs pour un dépôt de *NosoBase*

- Totaux minimums

Pour un nom (recherche d'antériorité + dépôt) :	2 875 F
---	---------

Pour 4 noms (1 recherche d'antériorité + 4 dépôts) :	8 125 F
--	---------

Tableau 12 : coût total pour la protection à l'INPI de *NosoBase*

3. 6. Avancée des travaux

En Septembre 1996, nous sommes rentrés dans le dispositif. Il a été décidé de ne protéger que le nom *NosoBase* pour l'instant, sans logo.

J'ai donc demandé une recherche d'antériorité, dont les résultats viennent d'être reçus et doivent être traités. En fonction de l'existence ou non d'une appellation similaire, une demande de dépôt sera envoyée.

4. Perspectives pour *NosoBase*

NosoBase n'en est qu'à ses débuts sur Internet. On ne peut pas savoir, bien sûr, quelle suite sera donnée à cette solution (des changements de personnel, de politique, de finances... peuvent se produire), mais dans une vision optimiste, il reste des tas de points à améliorer ou ajouter :

- Saisie par macros Word

répartir les revues entre les C.CLIN

Tester le programme et récupérer des notices saisies par d'autres C.CLIN

adapter une version pour Macintosh

le C.CLIN Paris-Nord souhaite créer une base de données sur sa documentation interne, et saisir les notices *NosoBase* sous Access : il faudra donc créer un autre programme de reformatage.

- Serveur WAIS/WWW

perfectionner comme prévu l'interrogation de la base (voir page 66), dès fin Septembre

faire la promotion du serveur sur Internet et dans le public du C.CLIN (voir page 76)

faire régulièrement la mise à jour (transfert des notices et réindexation)

intégrer un compteur de connexions sur le serveur

créer une liste de discussion NosoBase, privée (demande d'inscription auprès de Nathalie LEROY) mais non modérée

ajouter la possibilité de consulter le thesaurus pour utiliser correctement le champ *Mot(s)-Clé(s)*

améliorer l'homogénéité de la base (nom des revues identiques, rajouter les résumés en langue originale manquants...), résultat de la démarche qualité commencée en Septembre

créer un script pour saisir les notices à distance (voir page 77)

créer un script pour commander les articles visualisés

accroître le contenu : intégrer plus d'informations sur la base, l'équipe, mais aussi les bulletins d'informations sur les réseaux ISO, REA...

proposer des liens vers d'autres serveurs sur les infections nosocomiales (ceci suppose un gros travail de recherche en amont, et de surveillance en aval).

- Dépôt du nom

finir le traitement de la recherche d'antériorité (voir page 86) et déposer *NosoBase* si possible.

Il y a donc encore du travail : avant d'effectuer mon Service National en Novembre, je reste travailler au C.CLIN Sud-Est pour un CDD d'un mois en Octobre.

- Et, dans une plus large optique, mise en place d'une chaîne SGML pour gérer l'ensemble du système (voir page 25).

Conclusion

Cette aventure de quatre mois a porté NosoBase sur la vague Internet.

Non sans permettre à son équipage d'acquérir de la "bouteille", de l'expérience. En effet, les différents mécanismes de mise en place d'un système documentaire sur Internet m'ont apporté un savoir-faire, non négligeable pour mon avenir professionnel.

De la récupération des données, première étape, au serveur WWW, aboutissement final, une vision d'ensemble est toujours nécessaire. Les différentes étapes, mêmes lorsqu'elles apparaissent en cours de route, sont une nouvelle situation à appréhender, comprendre, surmonter et maîtriser. Et l'on peut dire qu'au long de ce stage, elles ont été nombreuses et fructueuses : mise en place d'une base WAIS, fonctionnement des scripts CGI, écriture HTML, création de formulaires, étude du format des données pour leur transformation, utilisation de langages de programmation, démarches de protection industrielle, etc... Sans compter une nouvelle expérience en entreprise, une autre approche du milieu documentaire et une découverte des métiers de la Santé. Je remercie à ce propos toutes les personnes du Pavillon 1 M qui m'auront aidées au cours de ce stage, et particulièrement Nathalie LEROY et Louis AYZAC.

Sans aucun doute, ce voyage valait donc la peine d'être tenté. Souhaitons à NosoBase un "cybervoyage" plein de bonnes surprises...

...et ils eurent beaucoup de notices bibliographiques.

Bibliographie

Les règles bibliographiques s'inspirent de la norme AFNOR Z44-005².

- **WAIS**

LINK-PEZET (Jo) et LACOMBE-CARRAUD (Elisabeth). - WAIS (Wide area information server). - In : Guide pratique d'Internet. - ed. A Jour, 1995, pp.80-99

CADIOU (Colette). - Conception d'un serveur WAIS pour une base bibliographique. - *Extrait de* : Conception de serveurs WAIS au CEMAGREF : étude de faisabilité, organisation, cahier des charges. - Rapport de stage DESS Informatique Documentaire. - ENSSIB, Université Lyon1, 1994

BARTHELEMY (Pierre) et ROLLAND (Robert). - WAIS [Wide Area Information Servers]. - In : L'internet professionnel. - ed. CNRS EDITIONS, 1995. - ISBN 2-271-05256. - Monographie n°1 du Micro Bulletin

- **HTML**

MAGID (Jonathan), MATTHEWS (R. Dougls) et JONES (Pau). - Serveurs web [outils et techniques]. - Paris, ed. International Thomson Publishing France, 1996 - ISBN 2-84180-052-0. - accompagné d'un CD-ROM

SAVOLA (Tom). - HTML. - Paris, ed. Simon & Schuster Macmillan, 1996. - ISBN 2-7440-0055-8. - accompagné d'un CD-ROM

- **BASIS PLUS / BASIS SERVER**

² AFNOR. - Norme française Z44-005 [Références bibliographiques]. - In : AFNOR, Recueil des normes françaises 1986 : Documentation [présentation des publications, translittération, thesaurus et indexation]. - tome 1, 3ème ed.. - Paris, Afnor, 1986, pp. 138-148. - Bibliothèque de l'ENSSIB : côte ENSSIB 025.324 AFN

LUBKOV (Michel). - Basis plus fait feu de tout bois. - *Archimag*, 90, déc.-
janv. 1996

• **REVUE HygièneS**

HygièneS [revue officielle de la Société Française d'Hygiène Hospitalière]. -
ISSN 1249-0075

Bookmark : adresses Internet

Le terme *bookmark* est repris du logiciel Netscape Navigator : il indique un répertoire d'adresses Internet. Comme partout sur Internet, les pages mentionnées sont susceptibles de disparaître : elles sont, en tous les cas, valables en Septembre 1996. Respectez bien les différences majuscules/minuscules.

- **Nosobase : serveur du C.CLIN Sud-Est :**

<http://www.univ-lyon1.fr/LyonSud/nosobase/departf.htm>

- **CISM**

page de garde :

<http://cismserveur.univ-lyon1.fr/CISM/>

bases WAIS gérées par le CISM :

http://www.univ-lyon1.fr/bases_de_donnees/entree.html

- **WAIS**

cours « en-ligne »

François Dagorn : Wide Area Information Servers

<http://www.univ-rennes1.fr/doc-html/sir/node66.html>

Pierre Barthelemy : le projet WAIS (Wide Area Information Servers)

<gopher://ns.urec.fr:70/00/Services/wais.docs/WAIS.Barthelemy.txt>

groupes de discussion (*newsgroup*) spécialisés :

comp.infosystems.wais (en anglais)

fr.comp.infosystemes (en français)

liste de discussion française :

wais-fr@univ-lyon1.fr

Cette liste permet aux administrateurs de s'entraider (ouverte aux responsables de centres de documentation et aux utilisateurs avertis)

INRIA : *directory of servers* français (base **WAIS** : interrogeable par client spécifique ou configuration de proxy WAIS sur un visualiseur HTML) :

Port : 210

Serveur : zenon.inria.fr

Nom de la base : directory-zenon-inria-fr

FAQ WAIS : une liste de questions fréquemment posées (pour les non initiés en général : FAQ, pour Frequently Asked Questions)

<http://www.cis.ohio-state.edu/hypertext/faq/usenet/wais-faq/freeWAIS-sf/faq.html>

FreeWAIS-sf

Université de Dortmund :

<http://ls6.informatik.uni-dortmund.de/ir/projects/freeWAIS-sf/>

WSC Group (reprend le manuel de Dortmund) :

<http://hecate.wsct.wsc.com/freeWAIS-sf/fwmain.html>

petit guide français :

<http://www-mathdoc.ujf-grenoble.fr/~cherhal/docs/freewais-sf.html>

SFgate

Université de Dortmund :

<http://ls6.informatik.uni-dortmund.de/ir/projects/SFgate/welcome.html>

FreeWAIS-sf et SFgate :

<http://www-fog.bio.unipd.it/waishelp/waishlp3.html>

Winwais 2.2A (logiciel WAIS client) :

<http://sir.univ-rennes1.fr/pub/PC/general/wais/winwais/>

- **CGI / Perl**

explication générale CGI :

<http://hoohoo.ncsa.uiuc.edu/cgi/intro.html>

ressources CGI sur Internet (en français) :

<http://www.edelweb.fr/EdelStuff/EdelPages/AdminW3/forms.html>

écritures de scripts CGI (en Perl) : travaux pratiques

<http://www.catt.ncsu.edu/users/bex/www/tutor/index.html>

Didacticiel Perl :

<http://agora.leeds.ac.uk/Perl/start.html>

- **FTP**

logiciel client WS_FTP :

ftp://ftp.grolier.fr/pub/simtelnet/win3/inet/ws_ftp.zip

le logiciel peut être récupéré sur de nombreux autres sites FTP

- **Visualiseurs ("browsers") Internet**

Netscape Navigator : site de Netscape

<http://home.netscape.com>

Internet Explorer : site de Microsoft

<http://home.microsoft.com>

- **Conception de serveur WWW**

Cours HTML : introduction

CNAM :

http://web.cnam.fr/cours_html/bienvenue.html

NCSA

<http://www.ncsa.uiuc.edu/General/Internet/WWW/HTMLPrimer.html#CharFormat>

Introduction HTML aux frames :

http://home.netscape.com/assist/net_sites/frames.html

bibliothèques d'images

icônes :

<http://www.ncsa.uiuc.edu/General/Internet/WWW/HTMLPrimer.html#CharFormat>

fonds :

<http://www.ncsa.uiuc.edu/SDG/Software/WinMosaic/Backgrnd/>

- **Adresses électroniques :**

mes adresses personnelles (jusqu'en Septembre / Octobre 1996) :

a l'ENSSIB

gutman@enssibhp.enssib.fr

au C.CLIN Sud-Est

jeromegu@uheim.uni-lyon1.fr

vous pourrez écrire à Nathalie LEROY ou Louis AYZAC (voir ci-dessous) **jusqu'au 1^{er} Janvier 1997** pour connaître ma nouvelle adresse

Nathalie LEROY

nathalle@uheim.uni-lyon1.fr

Louis AYZAC

louisay@uheim.uni-lyon1.fr

liste des élèves du DESS :

liste-dessid@enssib.fr

- **Bases de l'INPI (attention : la plupart sont sur Minitel)**

MARLYS (marques françaises depuis 10 ans) :

consultable dans un centre régional INPI

ICIMARQUES (marques) :

Minitel 36 29 36 30

EURIDILE (Registre National du Commerce) :

Minitel 36 17 EURIDILE

FMARK et TMINT :

sur abonnement (serveur Questel ?)

- **Divers**

Moteur de recherche Glimpse

<http://glimpse.cs.arizona.edu:1994/index.html>

ENSSIB

page de garde

<http://www.enssib.fr/>

DESS "Informatique Documentaire"

<http://enssibhp.enssib.fr/Enssib/dessid/DESSID.html>

Liste des abréviations utilisées

(reportez vous à l'index pour connaître les passages qui mentionnent ces abréviations)

ANSI	American National Standards Institute
BDSP	Banque de données Santé Publique
BIRD	Base d'Information Robert Debré
C.CLIN	Centres de Coordination de la Lutte contre les Infections Nosocomiales
CGI	Common Gateway Interface
CHLS	Centre Hospitalier Lyon Sud
CISM	Centre d'Information Scientifique et Médicale
CLIN	Comité de Lutte contre les Infections Nosocomiales
CNIDR	Clearinghouse for Networked Information Discovery and Retrieval
CTIN	Comité Technique National des Infections Nosocomiales
DRASS	Direction Régionale des Affaires Sanitaires et Sociales
DSII	Direction des Systèmes, de l'Information et de l'Informatique
ENSP	Ecole Nationale de Santé Publique
GED	Gestion Electronique de Documents
HCL	Hospices Civils de Lyon
HEIM	Hygiène, Epidémiologie et Information Médicale
HTML	Hyper Text Markup Language
HTTP	Hyper Text Transfert Protocol
INIST	Institut National de l'Information Scientifique et Technique
INPI	Institut Nationale de la Propriété Industrielle
ISO	International Organization for Standardization / réseau de surveillance des Infections du Site Opérateur
NSF	National Science Foundation
OCR	Optical Character Recognition
REA	réseau de surveillance des infections nosocomiales en REAnimation
ROCAD	Reseau Optique du CAmпус de la Doua
SGML	Standard Generalized Markup Language

WAIS	Wide Area Information Server
WWW	World Wide Web

Index

A

AFNOR 80
Apple Computers 42
AYZAC 13; 21; 24; 39; 40; 53; 54; 67

B

Basis plus 20; 21; 23; 24; 28; 29; 30; 31; 32;
33; 34; 37; 39; 40; 50; 51; 52; 53; 61; 71
Basis server 40
BDSP 18; 19
Bibliothèque de Médecine 13
BIRD 19
BOPI 74; 75
Bull 20

C

Cahier des charges 23
Caractères accentués 60
Cassettes-vidéo 16; 20
CD-ROM 11; 13; 14; 68; 73
Centre Français de la Copie 19
Centre Hospitalier Lyon Sud 10
CGI 48; 49; 50; 55; 69; 70; 71
CISM 40
CLIN 8
Clipper 53
CNIDR 42; 44; 45
CTIN 8
Current Contents 27

D

Digital Média 68; 71
Directory of servers 44; 45; 54
Documentation 11; 13; 15; 19; 20; 21; 22; 24;
41; 55; 59; 77; 82

Dow Jones 42
DRASS 19
Droits de copies 19
DSII 20; 24; 25; 31; 39

E

Editeurs HTML 62
Enquête Nationale de Prévalence 11
ENSP 19
ENSSIB 41
EPI INFO 14
EURIDILE 73
Expressions réservées 70

F

FABRY 11; 12; 14
Fichier *.dot* 35
Fichier *.fmt* 51; 52
Fichier *.htaccess* 71
Fichier *.src* 44; 46; 54
FMARK 73
Fonds documentaire 16
Formulaires 45; 49; 54; 55; 58; 66
Fourniture de documents 19
Frames 62
FTP 14; 29; 30; 38; 54; 68

G

GED 21
Glimpse 40
Grenoble 12

H

HCL 10; 20; 21; 23; 24; 29; 30; 31; 40; 71
HEIM 10; 13; 14; 23; 24; 30; 40
HygièneS 11; 20; 69

I

ICIMARQUES 73
ID France 20; 40
Index 42; 43; 46; 47; 49; 55; 58
INIST 19
INPI 71; 72; 73; 74; 75; 76
INRIA 42; 43; 44; 54
INSA 41
Internet Explorer 14; 26; 63
Interrogation avancée 55; 66
Interrogation simple 58; 66
Intranet 23; 68
ISO 11; 42; 78

L

LEROY 12; 15; 18; 21; 23; 30; 39; 54; 61; 67;
69; 77
Ligne d'en-tête 43; 46; 50; 54; 59; 60
Logiciel documentaire 20; 21; 24; 26

M

Macros 30; 31; 35; 36; 38; 53; 77
MARLYS 73
Medline 13; 14; 16; 27
Microsoft Foxpro 53
Minitel 18; 19; 73
Mistral 20; 31
MONNET 12
Moyens informatiques 13; 21
Mozilla 62

N

Netscape 14; 26; 62; 63; 64; 65; 67
Novell 13; 14; 21; 40
NSF 42

O

OCR 21
Opérateurs 47

P

Pascal 16; 27

Passerelles 45; 48; 54
Perl 29; 48; 49; 70; 71
Personnel 15
Produits documentaires 20
Profil documentaire 20
Public 15

Q

Qualité 8; 15; 21; 78

R

REA 11; 78
Registre National du Commerce et des Sociétés
73
Regular expressions 70
Relevance feedback 44
Réseau de lecture 16
Revue bibliographique 20
ROCAD 13; 55; 58

S

SAPHIR 19
SAVEY 12
SFgate 48
SGML 21

T

Thesaurus 34; 60; 78
Thinking Machines Corporation 42; 44
TMINT 73
TOXIBASE 19

U

Université de Dortmund 40; 43; 45; 83
Université de Médecine 10; 54

W

WAIS 41
Waisindex 43; 46
Waissearch 43
Waisserver 43; 46
Winwais 43; 44

WordBasic 30

Z

Z39.50 42

Annexes

Vous trouverez dans les pages qui suivent :

- un exemplaire de la fiche manuelle d'indexation (4 pages)
- une interrogation de NosoBase sur notre serveur HTTP (4 pages):
 - formulaire HTML avec la question :
Mot(s)-Clé(s) : SIDA ET MATERNITE
 - lignes d'en-têtes retournées
 - affichage des documents concernés
- les impressions des pages de notre serveur Internet (le tout est cependant beaucoup plus "joli" sur un écran d'ordinateur connecté, surtout 17" !)
<http://www.univ-lyon1.fr/LyonSud/nosobase/departf.htm> (les adresses apparaissent en haut à droite) :
 - barre de menus (frame du bas : *menubarf.htm*) : 1 page
 - page de garde (*bienvenf.htm*) : 2 pages
 - interrogation simple (*intero1f.htm*) : 1 page
 - interrogation avancée (*intero2f.htm*) : 1 page
 - guide (*guidenf.htm*) : 7 pages
 - présentation de la base et des C.CLIN (*contenuf.htm*) : 4 pages