HAL
open science

CONCEPTION DE LA BASE DE DONNEES SPIDER ET PREPARATION DE SA MISE EN LIGNE SUR LE SITE SANTE-PUB

Sandrine Perrin

To cite this version:

Sandrine Perrin. CONCEPTION DE LA BASE DE DONNEES SPIDER ET PREPARATION DE SA MISE EN LIGNE SUR LE SITE SANTE-PUB. domain_shs.info.docu. 2000. mem_00000078

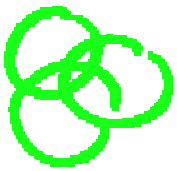
HAL Id: mem_00000078

https://memsic.ccsd.cnrs.fr/mem_00000078

Submitted on 17 Feb 2004

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

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



UNIVERSITE HENRI POINCARE, NANCY I
FACULTE DE MEDECINE
ECOLE DE SANTE PUBLIQUE

Mémoire de stage
Présenté dans le cadre du

DESS Information Scientifique et Technique
Promotion 1999-2000

CONCEPTION DE LA BASE DE DONNEES SPIDER
ET PREPARATION DE SA MISE EN LIGNE SUR LE SITE SANTE-PUB

LE RAPPORT

Stage effectué par **Mlle Sandrine Perrin**
Sous la responsabilité de :
Mme Martine Tabozzi et M. Etienne Rozé

DESS Cohabilité par trois universités

Remerciements

Travailler dans une ambiance détendue et joyeuse est le meilleur stimulant, je tiens à remercier le Professeur Jean-Pierre Deschamps, directeur de l'Ecole de Santé Publique de Nancy de m'avoir permis d'incorporer, pendant quelques mois, l'équipe du département "Epidémiologie, Sciences Humaines et Santé Publique".

Je remercie tout spécialement les deux responsables du projet, Mme Martine Tabozzi, maître toile du site de l'Ecole de Santé Publique, pour ses conseils et ses connaissances et M. Etienne Rozé, informaticien, pour sa disponibilité et pour m'avoir aider à percer les secrets de l'informatique.

Je tiens à exprimer ma reconnaissance à Mlle Sophie Osèle, documentaliste de l'Ecole, pour son dynamisme qui m'a accompagné au cours de nos après-midis de travail au Centre de Documentation.

Mes remerciements vont aussi à Philippe Legrand, mon co-stagiaire, pour m'avoir soutenu tout au long de ce projet.

Je ne saurais oublier Mlle Sylvie Giraudo pour sa bonne humeur, son soutien et l'aide précieuse apportée pour la rédaction de ce mémoire.

Je n'oublierai pas l'accueil chaleureux des membres de l'Ecole qui m'ont accueilli dans leur équipe pour ces quelques mois d'été.

Enfin je remercie tous les membres du jury pour avoir accepté de participer à ma soutenance et d'y apporter un regard critique.

RESUME

Dans le cadre du DESS IST (Informatique Scientifique et Technique) délivré conjointement par l'Université de Nancy I, l'Université de Nancy II et l'Institut Polytechnique de Lorraine, j'ai effectué mon stage pratique d'une durée de trois mois et demi à l'Ecole de Santé Publique de Nancy (ESP).

L'ESP a une double vocation d'enseignement et de recherche. Dans un souci d'améliorer ses services aux professeurs et aux étudiants, l'ESP a développé le projet SPIDER Santé-pub, il consiste à créer une base de données spécialisée en santé publique : SPIDER Santé Publique – Information et Documentation pour l'Enseignement et la Recherche qui est mis en ligne sur le site portail Santé-pub.

J'ai participé à la première partie de ce projet en réalisant la base de données SPIDER, née de la fusion de bases existantes au sein de l'Ecole et d'une grille d'évaluation de sites : "Santé publique", permettant de référencer les sites dans l'annuaire inclus dans SPIDER. Ces deux outils sont maintenant opérationnels et installés sur le site Santé-pub.

Pour la conception du site Santé-pub, j'ai travaillé sur l'ergonomie et la navigation du site web et j'ai conçu les formulaires et défini les modalités d'interrogation et de saisie en fonction des besoins des utilisateurs et des contraintes imposées par la documentaliste. Le site Santé-pub sera étoffé et diversifié au cours du temps pour devenir une référence dans la recherche documentaire en santé publique.

La participation à ce projet m'a fait découvrir le travail au sein d'une équipe pluridisciplinaire (documentation et informatique), à modéliser des besoins en suivant des contraintes techniques, humaines et financières. Les outils conçus m'ont permis de mieux connaître Internet et les technologies nécessaire à la conception d'un site et à sa mise en ligne. Parallèlement j'ai travaillé sur des aspects purement documentaires tels que la catalogage et l'indexation.

10 mots-clés

Site web, formulaire, grille d'évaluation, interrogation, logiciel documentaire, base de données, modélisation, santé publique, Internet, ergonomie.

SOMMAIRE

Introduction 6

I Contexte et projet

1.1 L'Ecole de Santé Publique : ESP 7
1.2 Le Centre de Documentation de l'ESP 8
1.3 Le projet base SPIDER et site Santé-pub de l'Ecole 9

II Structure de la base de données SPIDER : Santé Publique - Information et Documentation pour l'Enseignement et la Recherche

2.1 Les bases de données existantes 12

2.2 Le contenu de la base SPIDER 12

2.3 Définition des champs figurant dans SPIDER
2.3.1 Comparaison des notices obtenues avec les bases existantes 13
2.3.2 Etude des notices visualisées sur Internet 17
2.3.3 Liste des champs et des formats retenus 17
2.3.4 Les champs d'indexation 18

2.5 Annuaire de sites : grille d'évaluation des sites web
2.4.1 Critiques de grilles existantes 19
2.4.2 Méthode utilisée 20
2.4.3 Questionnaire et pondération des critères d'analyse 20
2.4.4 Validation de la grille et mise en place 21

2.6 Construction de la base
2.6.1 Matériel à disposition 22
2.6.2 Schéma entité association 23
2.6.2 Importation des données 23
2.6.3 Critiques 24

III Conception du site Santé-pub

3.1 Objectifs du site Santé-pub 25

3.2 Méthode de travail
3.2.1 Présentation de l'ergonomie du site 27
3.2.2 Le matériel employé pour la construction 28

3.3 Construction du site : les formulaires
3.3.1 Les types utilisateurs 30

3.3.2 Module d'interrogation
3.3.2.1 Les formulaires d'interrogation 32
3.3.2.2 Page réponse et page d'aide 33

3.3.3 Module de saisie
3.3.3.1 Pour tout public 34
3.3.3.2.Pour la documentaliste 35

3.3.4 Module bibliographie 35

3.4. Evolution du site 36

Conclusion **37**

Bibliographie **39**

Liste des sites visités **40**

Introduction

La santé publique, spécialité médicale, souffre depuis longtemps du manque d'intérêt qui lui est accordé. Toutefois cette discipline démontre de plus en plus son caractère essentiel dans l'élaboration des politiques publique de santé en France et dans le monde.

L'Ecole de Santé Publique de Nancy veut profiter de cet engouement afin de mieux faire connaître et reconnaître cette activité.

Le service technique de l'école a formé le vaste projet de concevoir un site spécialisé en santé publique fournissant différents services tels que la mise en ligne de la base de données **SPIDER (Santé Publique - Information et Documentation pour l'Enseignement et la Recherche)** proposant des "**références bibliographiques**" et un "**annuaire de sites en santé publique**", ainsi que des **forums de discussion** pour créer un réseau de professionnels et d'étudiants en santé.

Le but est que ce service devienne une référence dans la recherche d'informations spécialisées en santé publique grâce à la qualité et des informations fournies et des conseils donnés.

Le projet, présenté dans ce rapport, répond aux besoins de l'Ecole, à savoir le remplacement du logiciel documentaire du Centre de Documentation et la mise en ligne du fonds documentaire et des écrits des chercheurs de l'Ecole sur un site spécialisé en santé publique : Santé-pub.

Ce travail est divisé en deux parties : une partie "modélisation", qui m'est confiée et une partie "technique", qui est réalisée par un stagiaire en informatique.

Ma mission consiste à identifier les besoins immédiats de l'école en collaboration avec la documentaliste et les informaticiens, mais également à planifier leur concrétisation au sein de la base de données SPIDER et du site Santé-pub.

Après une présentation générale de l'Ecole de Santé Publique et de son projet, le travail réalisé est présenté en deux partie. Tout d'abord, les étapes de la **préparation de la base de données** et la **conception de la grille en santé publique** sont explicitées. Ensuite une maquette du site web est présentée, elle est axée essentiellement sur l'élaboration de **formulaires d'interrogation et de saisie.**

I Contexte et projet

1.1 L'Ecole de Santé Publique : ESP

Historiquement, le Professeur Jacques Parisot (1882-1967), pionnier de l'éducation pour la santé, de la prévention et de la réadaptation, a défini les concepts fondateurs de l'Ecole en 1930. Le laboratoire de Santé Publique de la faculté de Médecine de Nancy ne prit le nom d'Ecole de Santé Publique qu'en 1990.

L'Ecole de Santé Publique de Nancy (ESP) est organisée en deux départements : le département "**Epidémiologie, Sciences Humaines et Santé Publique**" sous la direction du Professeur Jean-Pierre Deschamps et le département "Santé et Environnement" sous la direction du Professeur Philippe Hartemann.

Elle propose des formations diplômantes ou qualifiantes, dans le domaine de la santé publique : licence, maîtrise, DEA, DESS, DU (diplôme universitaire) ou modules isolés.

Les principaux thèmes proposés sont les suivants :

- Santé et développement,
- Santé publique et communautaire,
- Politiques européennes de santé,
- Sciences sociales appliquée à la santé.

Ces formations sont accessibles aux étudiants, mais également aux professionnels de santé qui souhaitent s'inscrire en formation continue. Plus de 700 personnes par an suivent ces enseignements.

La particularité de l'ESP est de proposer un enseignement à distance (EAD). Il représente plus de la moitié des étudiants (environ 300 élèves répartis à travers le monde). L'objectif de l'école est de former du personnel compétent à un niveau local ou régional, pour participer à la réalisation de programmes ou à la mise en place de politiques de santé publique.

L'EAD pose certaines contraintes dans l'organisation des enseignements. La majorité des cours se fait à partir de fascicules envoyés par la poste (solution longue et coûteuse). La possibilité de suivre les enseignements par Internet est offerte aux étudiants le désirant. Par un accès réservé sur le site de l'ESP, les étudiants disposent des fascicules de cours, des plannings et d'autres services facilitant l'apprentissage. L'Ecole souhaiterait améliorer ses services "à distance".

Outre son activité d'enseignement, l'Ecole participe à des projets de recherche en collaboration avec plusieurs pays en Europe, en Afrique et en Asie du Sud-Est. Ses principaux domaines d'études sont : l'hygiène (travail sur l'eau et l'environnement, sur la lutte contre les infections hospitalières), la médecine sociale (travail sur la prévention et l'éducation pour la santé), l'épidémiologie et les sciences sociales.

La partie recherche des activités de l'Ecole est contrôlée par l'UPRES, l'Unité Propre de la Recherche et l'Enseignement du Supérieur.

1.2 Le Centre de Documentation de l'ESP

Pour fournir un cadre de travail performant aux étudiants et aux chercheurs, l'Ecole et plus précisément le département "Epidémiologie, Sciences Humaines et Santé Publique", possède son propre Centre de Documentation. Depuis deux ans, ce centre est rattaché à la Bibliothèque Universitaire de la Faculté de Médecine de Nancy.

Le fonds documentaire comprend :

- ✧ des ouvrages, des actes de congrès, des rapports (environ 5 000 références),
- ✧ environ 80 de titres de revue,
- ✧ les mémoires d'étude et de recherche réalisés par les étudiants dans le cadre des diplômes de l'ESP et ayant obtenus une note supérieure à 14,
- ✧ environ 200 dossiers thématiques alimentés régulièrement par la documentaliste grâce à une revue de presse.

Les principaux domaines d'informations traités par le fonds documentaire sont :

- ✧ santé dans les pays en développement (les publications de l'OMS …),
- ✧ systèmes de santé (en Europe et dans le monde) et protection sociale,
- ✧ sciences humaines appliquées à la santé
- ✧ santé maternelle et infantile, santé des adolescents, santé des personnes âgées,
- ✧ santé des groupes sociaux défavorisés,
- ✧ méthodes statistiques et épidémiologiques.

Ce fonds est actuellement géré avec **deux logiciels documentaires.**

1/ Le **logiciel GESBib** (version 3, datant de 1993) offre les fonctionnalités suivantes :

L'enregistrement d'une notice pour chaque document

Une notice comprend des données de **catalogage** (auteur, titre, année, éditeur…) et des données **d'indexation** et de localisation (liste de mots-clés ou autres informations précisant le contenu du document, cote de rangement). L'indexation se fait à partir du **thesaurus de la Banque de Données en Santé Publique (BDSP).**

La recherche de références

L'interface d'interrogation du catalogue est fonctionnelle mais peu conviviale, quasiment tous les champs de la notice sont interrogables. Toutefois les manipulations sont longues et peu intuitives.

La rédaction des bibliographies

Les références sont triées et présentées selon les **normes de Vancouver**. Ces normes de rédaction et de présentation des références bibliographiques, ont été établies par un groupe de recherche américain en 1978 et mis à jour régulièrement (cf. Annexe 6.1).

Le principal atout de GESBib réside dans ses **index**, à savoir une liste alphabétique de toutes les valeurs enregistrées d'un champ avec le nombre de documents y faisant référence. Les index sont nombreux et faciles à utiliser lors des saisies ou des recherches. Il est possible de les parcourir en tapant les premières lettres du mot recherché puis de valider son choix. Leur but est de limiter le plus possible la frappe.

Cependant ce logiciel est ancien et accessible seulement à partir de deux postes : un poste réservé à la consultation libre en salle de lecture de santé publique et un pour la documentaliste. Toutes les possibilités de GESBib sont loin d'être exploitées et celles utilisées sont similaires à celles offertes par tous les logiciels de gestion de base de données.

2/ Le **logiciel Vubis** est utilisé pour gérer la totalité du fonds documentaire de la Bibliothèque Universitaire de la Faculté de Médecine et, depuis deux ans, une partie du fonds du centre de documentation en santé publique (à terme toutes les notices devraient y figurer). Il fournit les mêmes fonctionnalités que GESBib.

Sa principale différence avec GESBib concerne l'indexation. Elle est réalisée ici à partir du thesaurus en médecine : **MeSH** (Medical Subject Headings) conçu par la National Library of Medicine (NLM-USA) , où la spécialité "santé publique" est très peu développée. Ce thesaurus est celui utilisé pour l'indexation des notices de **MedLine** : principale base de données bibliographiques en médecine produite par la NLM, qui couvre tous les domaines médicaux et connexes: biochimie, biologie, médecine clinique, économie, éthique, odontologie, pharmacologie, psychiatrie, santé publique, toxicologie, médecine vétérinaire.

➪ Afin de ne pas perdre la qualité d'interrogation du fonds documentaire en santé publique, la documentaliste conserve les deux logiciels (Vubis et GESBib).

1.3 Le projet base SPIDER et site Santé-pub de l'Ecole

Le projet de conception d'un site web découle des trois constats suivants :

- L'Ecole veut augmenter le nombre des services offerts aux étudiants à distance et surtout leur qualité,
- Le logiciel documentaire GESBib doit être remplacé par un logiciel mieux adapté aux besoins permettant, notamment, la mise en ligne du fonds,
- Les informations en santé publique présentes dans les nombreux sites en santé sont difficiles à extraire de la masse d'informations biomédicales, cliniques et thérapeutiques.

Pour débuter ce projet, l'ESP a décidé de mettre en ligne les **publications de son personnel**, **son fonds documentaire** et **un annuaire de sites en santé publique**. L'ensemble des données est stocké dans une seule base de données : **SPIDER** (Santé Publique - Information et Documentation pour l'Enseignement et la Recherche). Le nouveau site web Santé-pub sert d'interface à la base. Il assure les mêmes services que **GESBib** (enregistrement des notices, recherche de références et création de bibliographies).

Parallèlement, pour rendre le site plus dynamique et plus convivial, des **forums de discussion** sont proposés aux professionnels de santé et aux étudiants.

Ce site s'adresse à différents publics qui ont des besoins spécifiques. Les professionnels de santé et les acteurs de terrains sont plus intéressés par les publications alors que les étudiants se tournent plutôt vers les ouvrages de base. Pour fournir une réponse spécifique aux attentes, plusieurs **types d'utilisateurs** sont définis avec pour chacun des droits spécifiques.

Pour la réalisation du site, nous devons respecter certaines contraintes financières, techniques et humaines. Le matériel utilisé est identique à celui employé pour le site de l'école.

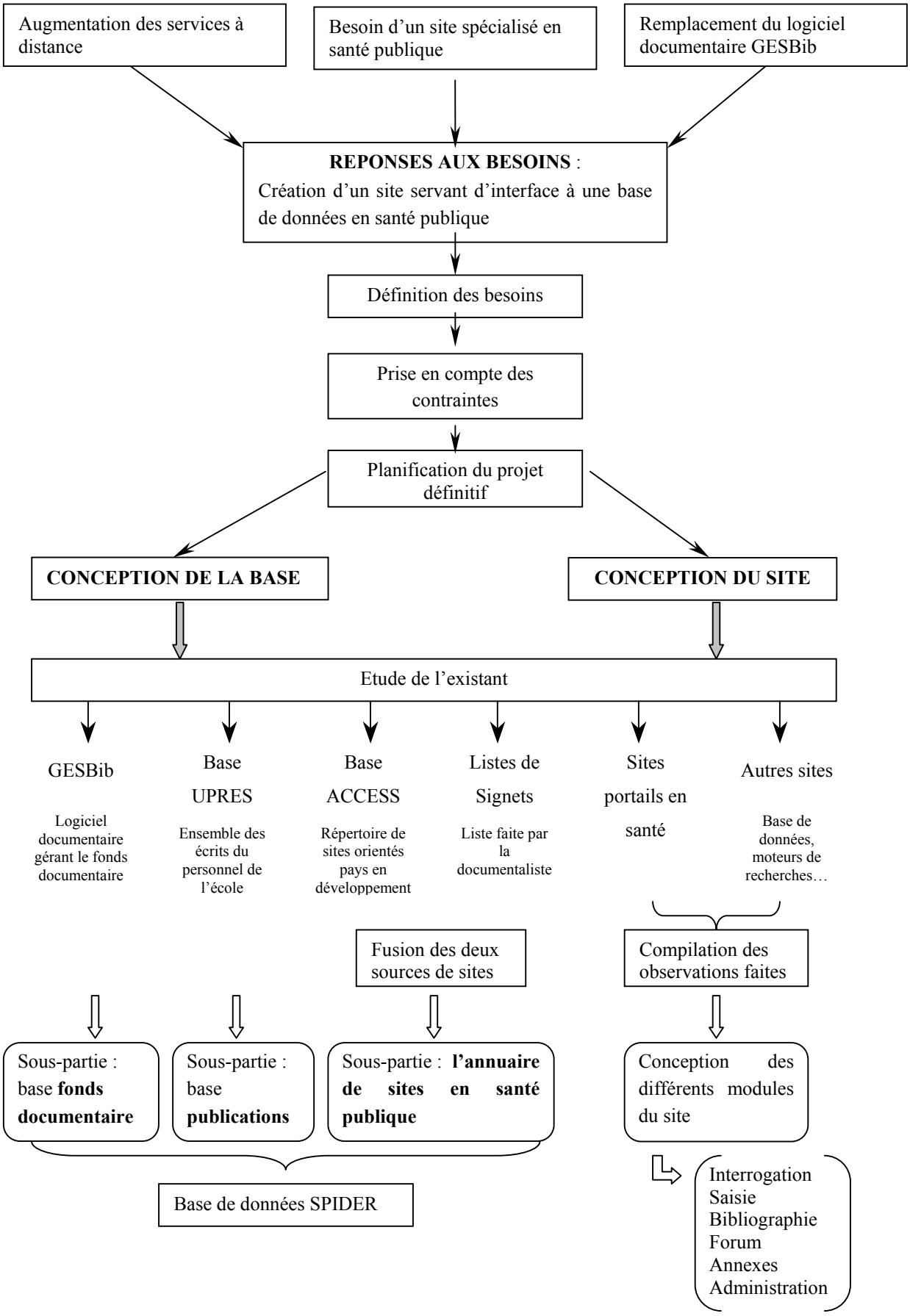
La construction et la maintenance du nouveau site s'ajoutent au travail de l'informaticien et du maître toile de l'Ecole (les administrateurs du site Santé-pub), ces activités doivent être au maximum optimisées grâce à des facilités d'utilisation et à l'ergonomie des interfaces.

Deux stagiaires débutent ce projet encadrés par l'informaticien et le maître toile :

- Ma mission consiste à étudier les besoins des utilisateurs potentiels et à prévoir le contenu de la base SPIDER et les pages du site Santé-pub. Le travail effectué est présenté dans le schéma n°1.
- La partie technique est réalisée par un stagiaire en informatique (DESS Informatique Double Compétence de Nancy). Suite à mon travail, il construit la base, l'alimente et commence à préparer le site (écriture des programmes pour traiter les requêtes sur la base, travail de mise en page des réponses).

L'achèvement du projet est à la charge des administrateurs du site.

SCHEMA n°1 : *PLAN DE TRAVAIL DU PROJET SPIDER ET SANTE-PUB*



II Structure de la base de données SPIDER : Santé Publique - Information et Documentation pour l'Enseignement et la Recherche

2.1 Les bases de données existantes

Le projet concerne quatre bases de données existantes au sein de l'Ecole, ayant des origines et des fonctions diverses.

La **base bibliographique de l'UPRES** (Unité Propre la Recherche et l'Enseignement Supérieur) est gérée par le logiciel Reference Manager. Elle contient les publications du personnel de l'Ecole : brochures, chapitres de livres, communications avec actes, communications avec publication, communications en attente, communications sans publication, livres, rapports, revues, revues à paraître, séries, thèses/mémoires et travaux non publiés.

Le logiciel Reference Manager est spécialisé dans la gestion de références bibliographiques. Son principal intérêt réside dans la sortie automatique de bibliographies respectant les syntaxes imposées par les éditeurs de l'ESP périodiques ou d'autres ouvrages.

Le **fonds documentaire** est maintenu par le logiciel documentaire **GESBib**, déjà présenté en détail dans le paragraphe 1.2. Les types de documents présents sont les livres, les congrès, les articles, les mémoires, les CD-ROM, les brochures et les fascicules.

Un **répertoire de site** a été conçu en 1999 sous **ACCESS** par un stagiaire lors d'un travail : l' "Information en santé publique sur Internet. Que trouve-t-on d'utile pour les administrateurs sanitaires africains ?". Il comprend une description détaillée d'environ 200 sites orientés vers les pays en développement. Les parties consacrées à l'indexation et à la critique des sites sont traitées de façon sommaire.

La deuxième source de sites exploités pour ce projet est le **fichier des signets** de la documentaliste, constitué au fur à mesure des recherches effectuées sur Internet. Les sites y sont simplement classés en grandes catégories.

2.2 Le contenu de la base SPIDER

La base SPIDER est alimentée à partir des quatre sources présentées précédemment.

La **base bibliographique de l'UPRES** est mise en ligne pour faciliter sa consultation. Actuellement, elle ne peut être consultée que par une seule personne à la fois (licence multiposte mais une seule base ouverte en même temps). L'objectif de la mise en ligne est de faire connaître le travail de recherche réalisé à l'Ecole.

La totalité de la base n'est pas importée dans SPIDER, seul les documents publiés le sont.

En raison des fonctionnalités spécifiques de Reference Manager dans l'édition de bibliographies, la base bibliographique de l'UPRES continuera à être gérée par ce logiciel. Une exportation périodique des nouvelles données sera effectuée pour actualiser SPIDER.

Le **fonds documentaire de l'Ecole** est importé dans son intégralité dans la nouvelle base. Le logiciel GESBib, quant à lui, est abandonné.

Les sites et les pages web référencés dans la **base de données ACCESS** devront être analysés afin de ne garder que ceux à importer dans SPIDER. Il s'agira essentiellement des notices décrivant des sites, plutôt que celles décrivant des pages web ou des documents en ligne axés sur un sujet trop pointu (les services sanitaires en Afrique).

L'ensemble des **signets** répertoriés par la documentaliste est intégré dans la base. Comme il ne s'agit que d'une liste d'adresses, les notices sur ces sites sont à créer en intégralité.

Une **alimentation périodique** de SPIDER se fera par importation de notices de la base bibliographique de l'UPRES ou par des saisies de la part de la documentaliste.

Afin d'enrichir la nouvelle base et de la rendre plus attrayante, la proposition d'ajouts de références par les visiteurs du site est possible uniquement pour les sites, les livres, les actes de congrès et les articles. L'ajout de mémoires et de thèses n'est pas possible, car il est difficile de vérifier les informations fournies. L'Ecole ne peut pas diffuser des informations erronées.

SPIDER regroupe des notices provenant de bases différentes. Ces origines diverses peuvent se retrouver lors de l'interrogation. Les internautes peuvent interroger

- soit la base dans son intégralité,
- soit les trois parties de la base,
  - les **publications** (base bibliographique et l'UPRES et ajouts depuis le site),
  - le **fonds documentaire** (logiciel documentaire GESBib),
  - l'**annuaire de sites** (fichier de signets, base sous ACCESS et ajouts depuis le site).

2.3 Définition des champs figurant dans SPIDER

2.3.1 Comparaison des notices obtenues avec les bases existantes

Après l'analyse des besoins, les champs devant figurer dans la nouvelle ont été sélectionnés. Il faut tenir compte :

- des champs présents dans les bases initiales,
- des informations nécessaires à la localisation d'un document,
- des informations donnant une idée du thème abordé pour aider les internautes dans leurs recherches.

La fusion de plusieurs bases implique également de compiler les champs de chacune de ces bases afin de définir ceux figurant dans SPIDER. Ce travail est parti de l'ensemble des champs présents dans la base bibliographique de l'UPRES et dans GESBib pour obtenir une seule et unique liste de champs. Ces étapes sont représentées dans le schéma n°2.

1/ Dresser la liste des champs

L'ensemble des champs existants pour chaque type de document et dans chaque base (UPRES et GESBib) a été répertorié. Au-delà de son intitulé, le contenu du champ est étudié, car des intitulés différents peuvent contenir des données identiques.

Ainsi 11 listes ont été obtenues : 4 pour la base bibliographique de l'UPRES (communication, livres, articles et mémoires) et 6 pour GESBib (congrès, livres, articles, mémoires, CD-ROM et divers – fascicules et brochures)

2/ Déterminer les champs nécessaires pour la nouvelle base

A l'intérieur de ces 10 listes, les champs jugés inutiles pour la base SPIDER ont été supprimés :

- les notices issues de l'UPRES vont incorporer la partie "Publications" de la nouvelle base. Seul les champs de catalogage et d'indexation ont été conservés.

- les notices issues de GESBib qui constituent la partie "Fonds documentaire" de la nouvelle base sont conservés en intégralité.
Après discussion avec la documentaliste, certains champs ont même été rajoutés pour affiner les informations proposées.

3/ Fusionner les listes par type de document

Parmi les 10 listes modifiées précédentes, certaines traitent du même type de document. Ainsi les listes doubles sur les congrès, les livres, les articles et les mémoires ont été fusionnées. Il reste 6 listes (cf. Annexe 1).

4/ Dresser une liste unique

La lecture des 6 listes obtenues montre que beaucoup de champs restent toujours en commun. Elles ont pu être regroupé en une seule liste qui a servi lors de la conception de la base.

5/ Etudier les formats des champs

Il est très important de comparer les formats de saisie de données, plus particulièrement pour les champs numériques et en particuliers ceux de format "Date" :

Par exemple : GESBib donne les dates avec le format suivant : JJ/MM/AAAA, et l'UPRES avec le format suivant MM/JJ/AAAA.

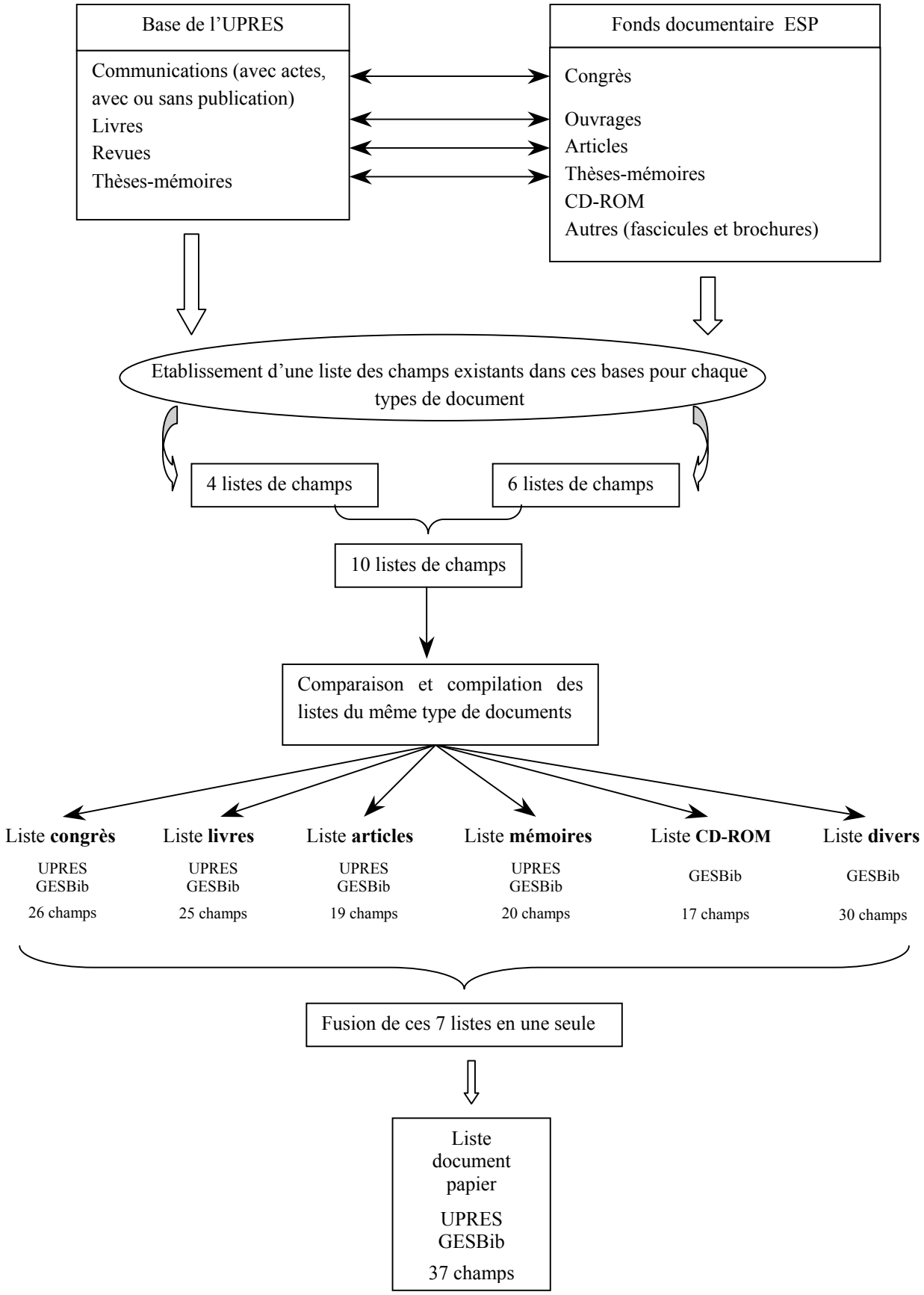
6/ Homogénéiser les formats

Il est donc essentiel pour les champs de SPIDER de choisir un format compatible avec ceux présents dans les deux bases d'origine. En effet si on tape la requête "12/05/1999", sans homogénéiser les formats, on obtient les réponses suivantes :

- GESBib donnerait les documents datés du 12 mai 1999 et pour la base bibliographique de l'UPRES, ceux datés du 5 décembre 1999.

Dans le cas du champ "Date", il faut choisir un format commun à toutes les notices, ici "JJ/MM/AAAA". Les champs qui possèdent un autre format devront être modifiés au cours de l'importation des notices dans SPIDER (cf. Annexe 2.1).

SCHEMA n°2 : *EXTRACTION DES CHAMPS FIGURANT DANS SPIDER POUR LES DOCUMENTS PAPIER*



2.3.2 Etude des notices visualisées sur Internet

Le choix des champs figurant dans l'annuaire de sites était libre de toutes contraintes. La base ACCESS de départ ne contient que le strict minimum d'informations.

La liste des champs décrivant les sites a été définie à partir des exemples trouvés sur Internet.

- ✧ visite de plusieurs sites : annuaire de site (essentiellement ceux présents dans les sites portails en santé), moteurs de recherche et base de données en ligne,
- ✧ analyse de la présentation des pages réponses à une requête,
- ✧ extraction des points positifs et négatifs,
- ✧ obtention d'une liste de 25 champs permettant de décrire l'origine et le contenu d'un site (cf. Annexe 1)

L'analyse des sites s'oriente dans deux sens : définition du sujet abordé et validation du contenu. Ce dernier aspect est de plus en plus traité dans les sites, et particulièrement en santé.

Pour valider le contenu, le moyen retenu consiste à utiliser une grille d'évaluation. Cette grille permet de juger plusieurs aspects d'un site grâce à un questionnaire. Elle est présentée dans le paragraphe 2.4 (p14).

2.3.3 Liste des champs et des formats retenus

A partir des listes de champs obtenus précédemment j'ai fait plusieurs tableaux sont faits pour préparer la construction de la base SPIDER. Ces tableaux sont proposés en annexe.

- ➢ tableau récapitulatif des champs présents en fonction du type de document (sites, congrès, livres, articles, mémoires, CD-ROM et divers) et de la base d'origine (base bibliographique de l'UPRES et logiciel documentaire GESBib). La présentation est faite de manière à mettre en évidence les champs communs entre les listes de champs. A partir de ce tableau j'ai décidé d'établir deux entités dans SPIDER : une entité pour les documents de type papier (congrès, livres, articles, mémoires, CD-ROM et divers) et une pour les sites (cf. Annexe 1.1).

- ➢ tableaux proposant pour chaque champ son contenu et l'origine de l'information enregistrée dans la base : un pour les documents papier et un pour les sites (cf. Annexe 1.2).

- ➢ tableaux indiquant la taille, le nombre de valeurs entrées et les formats possibles pour chaque champ. A partir de ces tableaux j'ai pu déterminer quels étaient les champs mono et plurivalents ainsi que les formats définitifs pour la base SPIDER (cf. Annexe 1.3 et 1.4).

- ➢ les derniers tableaux ont été écrits pour aider le concepteur du site et des formulaires de saisie et de recherche. Ils permettent de filtrer les champs présents dans les différentes pages : question, réponse ou saisie, mais aussi en fonction du format d'affichage demandé (cf. 3.3.2.1, paragraphe sur la Norme de Vancouver) (cf. Annexe 6.2).

2.3.4 Les champs d'indexation

L'indexation est un processus de lecture, de compréhension et de traduction du contenu d'un document dans un langage documentaire. Indexer un document consiste à lui attribuer plusieurs mots décrivant son contenu. Trois méthodes d'indexation différentes ont été rencontrées au cours de ce projet :

✧ **indexation à partir du langage libre** (les mots sont choisis au hasard). Cette méthode est employée dans le répertoire de sites sous ACCESS.

✧ **indexation à partir des mots d'une "liste d'autorité",** le choix des mots se limite à une liste définie par avance. Les documents de la base bibliographique de l'UPRES sont indexés avec une liste 130 mots.

✧ **indexation à partir d'un thesaurus**. C'est une liste de concepts structurée hiérarchiquement, relative à un ou plusieurs domaines de la connaissance ne comportant ni synonymes (mots différents mais possédant le même sens) ni polysémies (mots possédant plusieurs sens). Aucun mot extérieur ne peut être ajouté. Le thesaurus de la BDSP est utilisé par la documentaliste.

Cette étape est importante lors de la rédaction d'une notice et demande une certaine pratique pour obtenir une indexation cohérente d'un document à un autre.

Une mauvaise indexation produit soit du bruit (renvoi de réponses sans rapport avec la question), soit du silence (absence de réponses pertinentes). Ces deux phénomènes doivent être minimiser.

La méthode d'indexation retenue pour la base SPIDER est de conserver le thesaurus en santé publique. Ce choix est le plus pertinent et le moins long à mettre en place.

Les mots-clés présents dans les notices de la base bibliographique de l'UPRES et dans la base ACCESS peuvent être traduit en concept du thesaurus en santé publique.

Parallèlement pour optimiser les recherches, **deux niveaux d'indexation** ont été adoptés.

- **Avec les mots-clés du thesaurus de la BDSP** : autant de mots-clés que nécessaire peuvent être attribués à chaque document. C'est une indexation fine pour effectuer des recherches sur des sujets pointus et obtenir des réponses ciblées.

- **Avec des catégories**, c'est une liste de mots représentant les grandes thématiques qui sont traitées en santé publique, elle doit être par la documentaliste en collaboration avec des enseignants de l'école. C'est une indexation grossière pour obtenir une réponse exhaustive à un thème.

L'utilisation des catégories permet de classer l'ensemble des documents disponibles dans le centre de documentation : les références du fonds documentaire, les dossiers documentaires, les sites Internet.

L'indexation est un exercice d'apparence facile mais il nécessite de la pratique et beaucoup de rigueur pour assurer une bonne homogénéité et obtenir des réponses aux requêtes aussi exhaustives que possible. En effet, un même concept ne doit couvrir ni trop, ni trop peu de documents. Parallèlement il faut maîtriser l'utilisation d'un thesaurus. Pour l'instant, seule la documentaliste est véritablement rompue à ce type d'exercice.

2.4 Annuaire de sites : grille d'évaluation des sites web

2.4.1 Critique de grilles existantes

Une recherche sur Internet m'a permis de trouver plusieurs grilles d'évaluation de sites dont certaines conçues spécialement pour les sites en santé. En effet plusieurs organismes de santé (français, anglais ou canadien) ont créé des grilles d'évaluation de l'information présente dans un site. Elles servent à pallier l'absence de contrôle sur le contenu des textes diffusés sur le web.

Une grille se présente généralement sous la forme d'un **questionnaire à choix multiples** où chaque réponse représente des points, le total de ces points donne la **note finale du site**. Elle peut être donnée sous forme brute ou avec des icônes comme des étoiles. (cf Annexe 3.1)

Deux sites m'ont été particulièrement utiles :

- **La grille de sélection canadienne** expose clairement sa méthode d'évaluation. Pour chacune des 26 questions de la grille il n'y a que trois réponses possibles : faible, bien, très bien. La valeur de chacune de ces réponses varie selon le critère évalué : le contenu est le plus important (noté sur 60), puis la navigation (noté sur 20) puis la présentation visuelle (noté sur 14) et enfin l'accessibilité (noté sur 6), soit une note sur 100 .

- **Net Scoring** ("grille d'évaluation" en anglais) est une grille française utilisée par plusieurs sites proposant un annuaire de sites (Cismef, Webmedguide…). Ces sites proposent une présentation de la grille. Par ailleurs un site est dédié au Net Scoring. Il donne accès à des démonstrations du questionnaire pour tout public ou bien pour les professionnels. Les réponses proposées sont propres à chaque question.

L'étude de ces grilles avec les questions posées et la méthode de calcul de la note m'ont aidé à établir la grille utilisée dans la base SPIDER.

Les livres sur la conception de site web (cf. Bibliographie) sont une autre source d'inspiration pour le questionnaire.

2.4.2 Méthode utilisée

La principale difficulté dans l'élaboration d'une grille est de **limiter la subjectivité individuelle** pour que la note attribuée soit représentative de la valeur du site. L'accès au détail de la construction de la note donne une idée des points faibles et des points forts observés dans ce site.

La grille est conçue pour une utilisation simple afin d'évaluer correctement les sites :

- ✧ un vocabulaire précis destiné à restreindre les interprétations,
- ✧ une analyse orientée dans trois directions : le contenu, la navigation et la présentation graphique,
- ✧ un système de cases à cocher pour chaque question permet d'attribuer des points en fonction des différents critères d'évaluation, à chaque case correspond une valeur,
- ✧ une pondération faite de manière à privilégier le "contenu", afin d'accorder une prépondérance à la pertinence de l'information plutôt qu'à la présentation,
- ✧ des icônes illustrant le score obtenu par le site analysé.
- ✧ un guide d'accompagnement du questionnaire qui précise, pour chaque question, les paramètres à prendre en compte et dans quelle proportion.

2.4.3 Questionnaire et pondération des critères d'analyse

La grille en santé publique comprend quatre thèmes d'analyses :

☙ *Le contenu :*

La valeur des informations proposées est étudiée en fonction des dates, de leur origine, des concepteurs du site, ainsi que des règles de présentation des informations. Cette partie comprend 14 questions notées sur 110.

☙ *La navigation* :

Elle correspond à la facilité de consultation d'un texte et du passage d'un texte à un autre est un critère important pour exploiter au mieux les possibilités qui sont offertes par les pages web. Elle regroupe 10 questions notées sur 50.

☙ *La présentation graphique* :

Cet aspect varie beaucoup d'une personne à une autre. C'est pourquoi cette partie ne comprend que quatre questions notées sur 20, qui ne s'intéressent qu'à des critères essentiels pour la bonne compréhension des textes.

☙ *Les critères automatiques* :

Parallèlement aux critères d'analyses appréciés par l'œil humain, il est possible de récupérer des informations de manière automatique. Un programme informatique conçu par l'école analyse le texte source des pages HTML du site étudié. Il repère les balises données en paramètre et récupère les éléments recherchés :

- ✧ le nombre de liens : externes, intrapages, interpages ou non valide,
- ✧ l'utilisation de certains langages : java, php, cgi-bin,
- ✧ la taille du site,
- ✧ la récupération des métadonnées,
- ✧ la présence du site dans les moteurs de recherche,
- ✧ la nombre de citation dans les autres sites.

Ces critères sont notés sur 20, soit une note finale sur 200 (cf. Annexe 3.2).

☙ Le système de notation choisi pour la grille en santé publique s'inspire à la fois de celui observée dans Net Scoring et dans la grille de sélection canadienne. Pour chacune des 32 questions préparées, une liste de réponses est proposée avec une valeur qui est fonction du critère d'analyse : le contenu a le poids le plus lourd, puis la navigation et en dernier le design.

Les valeurs obtenues par ce programme sont stockées dans SPIDER. Périodiquement le serveur analyse les sites référencés dans la base pour actualiser les valeurs de ces critères automatiques. Au cas où les nouvelles valeurs seraient très différentes de celles enregistrées (par exemple : modification de plus ou moins 5% du nombre de liens), un message est envoyé à l'administrateur afin qu'il revérifie le contenu du site et décide son maintien dans la base ou non.

2.4.4 Validation de la grille et mise en place

La dernière étape de la conception de la grille en santé publique consiste à valider le questionnaire et la pondération adoptés en vérifiant leur cohérence.

Après avoir testé quelques sites avec la grille en santé publique, la grille canadienne et Net Scoring, les trois scores obtenus pour chaque site ont été comparés. Etant donné qu'ils tendaient tous vers le même avis, la grille en santé publique est jugée exploitable pour alimenter la base de données.

La grille de santé publique présente une particularité par rapport à celles étudiées sur le web : la présence de **critères automatiques**. La réponse obtenue pour ceux-ci est difficilement quantifiable. Pour apprécier la signification du nombre de liens par rapport au texte, il faut utiliser un échantillon de sites plus important. Leur exploitation demande un travail qui n'a pas été réalisé dans le cadre du stage.

Ces critères sont utilisés pour l'instant sous forme brute et à titre indicatif sauf pour quatre d'entre eux.

- ✧ La présence de liens non valides indique un manque de maintenance du site de la part du concepteur, elle entraîne une diminution de la note.
- ✧ L'absence de liens externes dénote d'un manque d'ouverture, elle entraîne aussi une diminution de la note finale.
- ✧ Le nombre de moteurs de recherche et le nombre de sites où le site étudié est référencé sont aussi évalués.

L'évaluation d'un site est une activité gourmande en temps et qui demande une certaine habitude. Elle est réalisée par les membres du personnel de l'école. Les sites sont répartis en fonction des spécialités de chacun. Pour les aider dans leur analyse, un guide d'accompagnement de la grille en santé publique a été rédigé (cf. Annexe 3.3).

La rédaction de la notice et le questionnaire vont de pair. Toutefois la partie de la notice traitant de l'indexation sera toujours réalisée par les mêmes personnes, qui doivent avoir l'habitude de manipuler un thesaurus et avoir déjà fait de l'indexation. L'indexation des sites se fait de la même manière que pour les documents papier de la base SPIDER (cf. 2.3.4), avec deux niveaux d'indexation : le choix d'une catégorie et le choix de plusieurs mots-clés issus du thesaurus de la BDSP.

2.5 Construction de la base

2.5.1 Matériel à disposition

La partie technique de la construction de la base a été faite par le stagiaire en informatique.

Les bases de données de l'école sont conçues avec deux logiciels :

- Sous **ACCESS** pour les bases spécifiques à chaque utilisateur. ACCESS est un système appelé SGBD pour Système de Gestion de Base de Données. Il propose un ensemble de services (applications logicielles) permettant de gérer les bases de données. Ce logiciel est développé par Windows.

- Le logiciel **MySQL** est utilisé pour les bases communes à tout le personnel ; par exemple la base "Agenda" propose le planning des professeurs de l'école avec une interface web.

MySQL est aussi un SGBD (Système de Gestion de Base de Données) installé sur le serveur web sous le système d'exploitation Linux. C'est un serveur de base de données SQL robuste. Le SQL ou Structured Query Language ou Langage de Requêtes Structuré est un des langages d'interrogation de base de données le plus utilisé actuellement. Il a été conçu pour supporter des bases volumineuses. Ses principaux avantages sont la rapidité, la robustesse et la facilité d'utilisation.

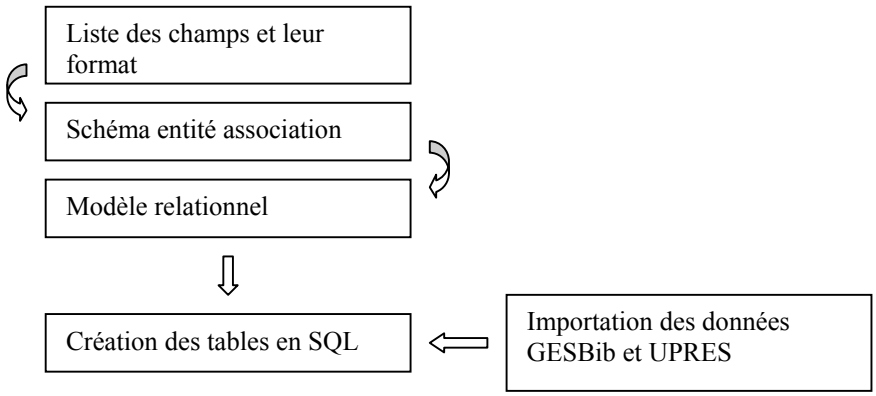
2.5.2 Schéma entité association

La base est créée sous le logiciel MySQL. Les principales entités sont les suivantes :

- ✧ une entité pour les documents papier avec tous les champs monovalents (avec une seule valeur),
- ✧ une entité pour les sites avec tous les champs monovalents,
- ✧ une entité par champs plurivalents (avec plusieurs valeurs) : mots-clés, auteurs…
- ✧ une entité pour les champs spécifiques à GESBib monovalents (ex : numéro de fiche, localisation, cote de rangement…

A partir de ces entités le modèle entité-association puis le modèle relationnel ont été construits. Ils sont présentés dans les annexes 2.2 et 2.3.

SCHEMA n°3 : *CREATION DE LA BASE DE DONNEES SPIDER ET IMPORTATION DES DONNEES*



2.5.3 Importation des données

L'importation des données consiste à prendre les données dans une base et les mettre dans une autre. Cette manipulation nécessite d'enregistrer dans des fichiers texte où des modifications pourront être apportées dans les notices.

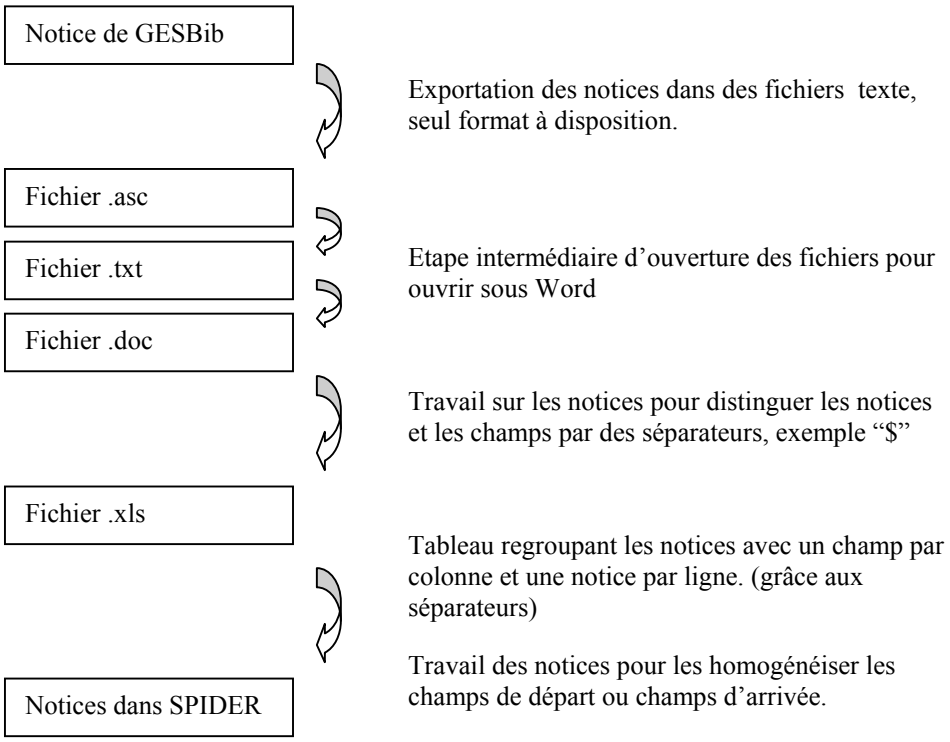
Dans le cas présent, deux importations ont été réalisées : une pour une partie de la base bibliographique de l'UPRES et une pour le fonds documentation de GESBib. Au cours de ce transfert, il a fallu changer le format de certaines dates pour homogénéiser.

Les différentes étapes nécessaires à l'importation des données depuis GESBib sont présentées dans le schéma 4.

2.5.4 Critiques

La structure du modèle relationnel assure une souplesse à la base permettant d'ajouter un nouveau champ dans une table ou ajouter une nouvelle table dans la base. La taille de la base est assez facilement gérable et l'interface est très pratique à utiliser.

SCHEMA n°4 : *LES DIFFERENTES ETAPES DE L'IMPORTATION DES NOTICES DE GESBIB VERS SPIDER*



III Conception du site Santé-pub

3.1 Objectif du site Santé-pub

Santé-pub a deux objectifs principaux :

✧ Etre une interface web pour la base SPIDER, en assurant l'interrogation et la saisie de nouvelles notices, par l'intermédiaire de formulaires.

✧ Etre un lieu de discussion pour les professionnels et les étudiants en santé publique grâce à la mise en place de forums.

Pour définir le contenu du site, une étude de quelques sites en santé a été menée, elle s'est orientée essentiellement vers les sites portails. Ces visites m'ont permis de dresser une liste de différentes observations. L'ensemble des informations récoltées est réparti en quatre groupes :

✧ Les informations présentes sur la page d'accueil : les rubriques proposées, les liens, la présentation des liens, des informations générales sur le site (date, origine…),

✧ Les formulaires d'interrogation : les types de recherche, les options d'interrogation, les fonctions annexes et la présentation des requêtes,

✧ Les réponses : la présentation des réponses, le contenu, le score et la mesure de la pertinence des réponses, les fonctions annexes,

✧ Divers : formulaires de contact, de saisie, de vote.

Pour chaque élément de chaque groupe, il faut déterminer son intérêt vis à vis du projet de l'école et sa faisabilité technique (cf. Annexe 4.1).

A partir de ce travail, les pages de Santé-pub sont définies. En fonction des objectifs définis pour le site, il est possible de les regrouper en six modules. Ils sont présentés dans le schéma n°5.

❶ *Module d'interrogation*

Il comprend des formulaires de recherche simple et avancée, et les pages réponses, ainsi que la page d'aide à la recherche.

❷ *Module de saisie*

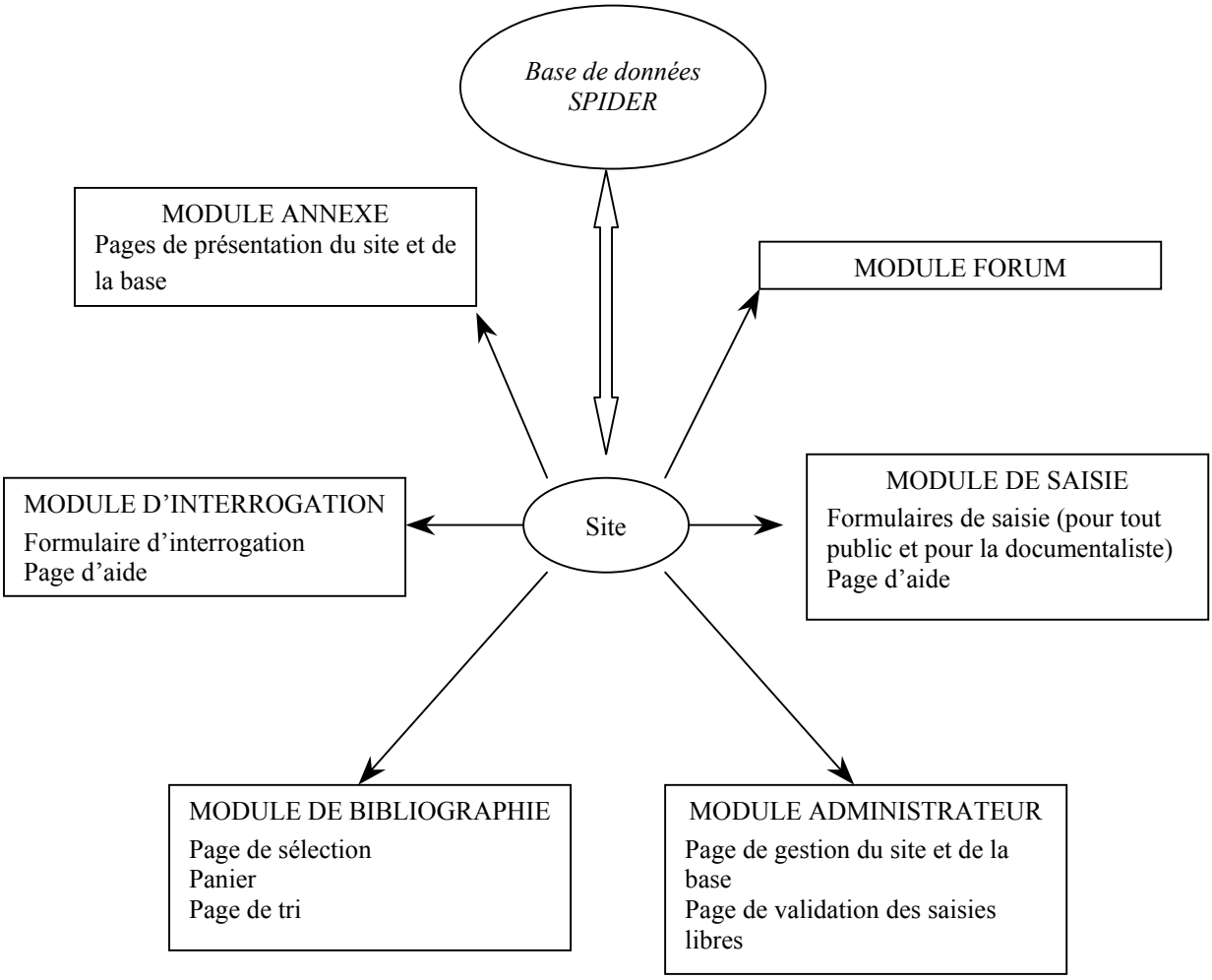
La saisie est distincte selon le type utilisateur.

Pour ajouter un document au fonds documentaire du centre de documentation, il n'existe qu'un seul formulaire pour tous les types de documents.

Pour la proposition d'ajout, par les visiteurs du site Santé-pub, un formulaire par type de document est créé : livres, articles, congrès, site.

Là aussi une page d'assistance à la saisie a été rédigée.

SCHEMA n°5 : *LES DIFFERENTS MODULES DU SITE SANTE-PUB*



❸ *Module bibliographique*

Il regroupe les pages permettant de sélectionner et de trier les références obtenues après une ou plusieurs interrogations de SPIDER. Ces manipulations permettent d'établir avec ces références une bibliographie normalisée avec la Norme de Vancouver. Cette bibliographie peut être imprimée ou enregistrer sous forme de fichier texte pour pouvoir ajouter des notices d'autres bases de données.

❹ *Module forum*

Les forums sont installés grâce à un programme appelé "Phorum" disponible sur Internet. Ce sont des pages paramétrables en langage php (cf. 3.2). Son installation est rapide.

Plusieurs forum sont créés dans Santé-pub : offres et demandes d'emplois, actualité… L'accès aux forums se fait par le rappel des principaux thèmes sur la page d'accueil.

❺ *Module annexes*

Ce module englobe des pages disparates servant à présenter le site, son contenu, ses objectifs, ses auteurs et également le plan du site.

❻ *Module administration*

Le site donne accès aux pages de gestion de la base et de gestion du site pour assurer sa maintenance et son suivi, mais aussi toutes les pages de validation des propositions faites depuis le site et l'accès à la grille en santé publique.

3.2 Méthode de travail

3.2.1 Présentation de l'ergonomie du site

Deux aspects sont à prendre en compte lors de la construction d'un site :

- ✧ le contenu des pages, ce travail a été présenté dans le paragraphe 3.1. Il se base sur les tableaux faits à partir des listes de champs pour établir les différents formulaires présentés ultérieurement.
- ✧ l'ergonomie du site, l'ensemble des liens qui relient les pages entre elles.

Pour mieux appréhender l'ergonomie, il est intéressant d'avoir une vue générale du site avec l'ensemble des pages et des liens.

Une représentation graphique du site est dessinée sous Word, elle comprend toutes les pages de Santé-pub, soit 40 pages et une partie des liens interpages : 40 liens aller-retour, soit 80 boutons (cf. Annexe 4.2). Le nombre total de boutons est beaucoup plus conséquent. Tous les liens ne sont pas présents dans le graphique pour une question de visibilité, il manque :

- ✧ tous liens de retour vers la page d'accueil présents sur chaque page du site,
- ✧ tous les liens partant du plan du site qui doit présenter toutes les pages présentes,
- ✧ les liens vers les index,
- ✧ les multiples liens intrapages entre les formulaires du site et les pages d'aide correspondantes.

Ce graphique a permis de mieux préciser les liens entre les pages. Toutefois il n'est pas suffisant pour vraiment tester la navigation. Ainsi une maquette du site Santé-pub est faite sous le logiciel Powerpoint (cf. Annexe 4.3). Ce logiciel sert à créer des diapositives et permet de les relier par des liens.

Le principe adopté est simple à chaque diapositive correspond une page du site. Les diapositives présentent un résumé de la page et l'ensemble des liens interpages.

3.2.2 Le matériel employé pour la construction

Le site est conçu avec le logiciel Webexpert, un logiciel shareware disponible sur Internet. Il a été choisi entre plusieurs en raison de sa facilité d'utilisation et l'étendue de ses capacités.

Deux types de pages sont présents dans le site : des pages HTML statiques et des pages dites dynamiques.

Les **pages HTML** statiques sont les pages de textes telles que les pages d'aide à la recherche et les formulaires.

Un formulaire est une page ou des informations peuvent être écrites grâce à des cases ou en cliquant sur des boutons, les informations ainsi notées sont envoyés à une adresse mail ou à un serveur web. Dans le site Santé-pub, ils servent à écrire des requêtes ou à saisir de nouvelles notices.

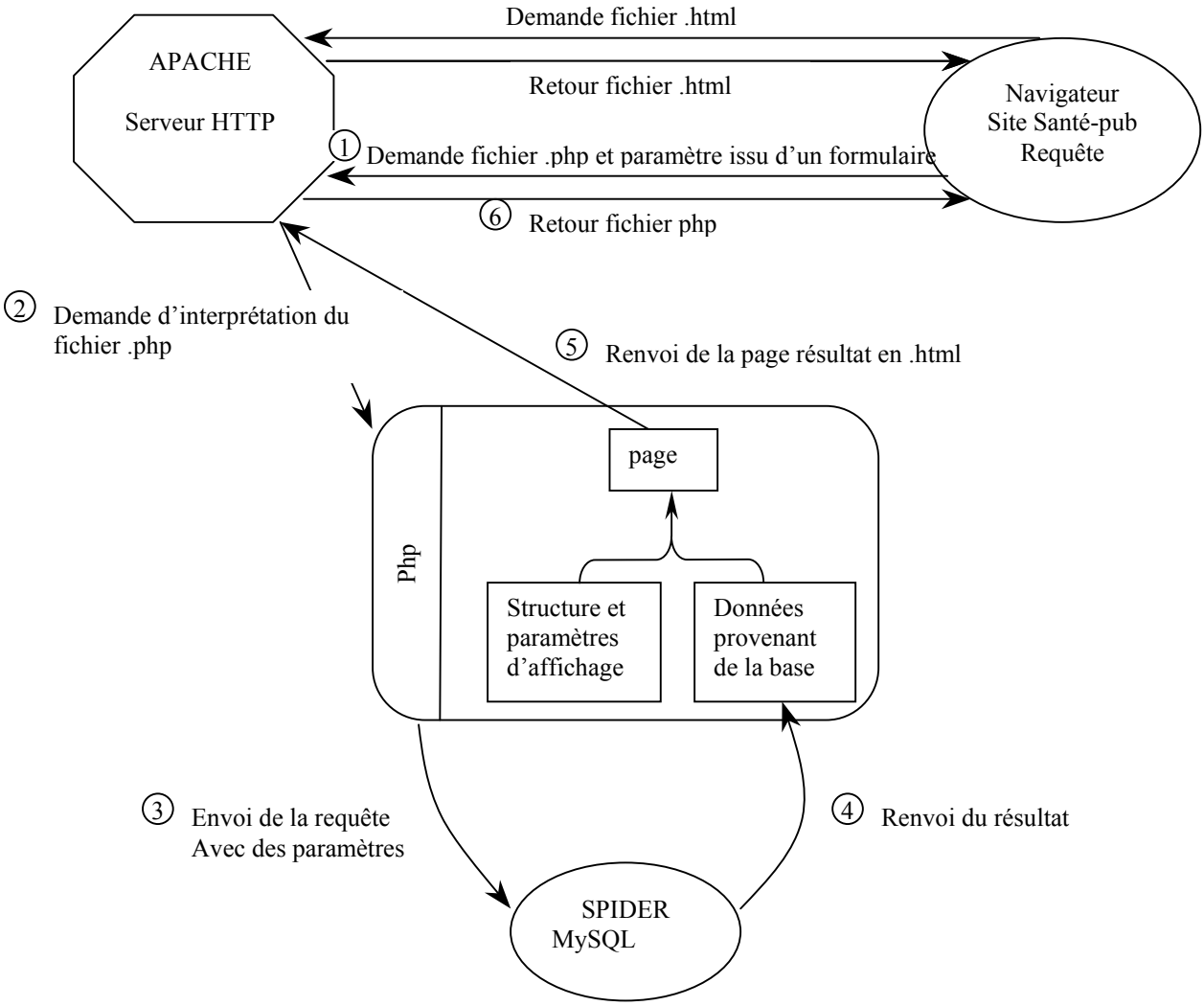
Les modalités de saisie sont variées :

- ✧ cases à cocher : sélection de une à toutes les propositions faites,
- ✧ bouton radio : sélection d'une seule proposition à la fois,
- ✧ zone de texte multilignes,
- ✧ zone de texte simple ligne,
- ✧ zone de texte déroulante,
- ✧ champs mot de passe,
- ✧ menu déroulant : un seul choix,
- ✧ liste déroulante : plusieurs choix possibles,
- ✧ bouton reset,
- ✧ boutons de commandes.

Les **pages "dynamiques"** sont faites en **langage php**. Ce langage fusionne un programme contenant la structure de la page et d'un programme contenant les données provenant d'un base de données, par exemple. La structure de la page peut être paramétrée par des indications notées dans un formulaire ou en fonction de restriction que le concepteur du site a lui même défini, par exemple en fonction du numéro IP, un certain nombre de champs de la base ne sont pas accessible. Un logiciel php fait la fusion des deux et assure la conversion du programme php en fichier html.

Le schéma n°6 montre le trajet suivi par les informations saisies dans un formulaire et le retour de la réponse vers le navigateur.

SCHEMA n°6 : *SCHEMATISATION DE LA CREATION D'UNE PAGE DYNAMIQUE PHP SUITE A UNE REQUËTE.*



3.3 Construction du site : les formulaires

3.3.1 Types d'utilisateurs

Le site est dédié à un public composé de professionnels de la santé, d'acteurs de terrain, d'étudiants ou de toutes personnes s'intéressant au domaine. Cependant toutes les informations mises en ligne ne sont pas accessibles pour tous.

La base est divisée en trois sous-parties : les "publications", le "fonds documentaire" de l'école et un "annuaire de sites". Mais chacune s'adresse plus particulièrement à un public ; les professionnels de santé et les acteurs de terrains sont plus intéressés par les publications alors que les étudiants se tournent plutôt vers les ouvrages de base.

Ainsi différents types d'utilisateurs ont été définis, avec chacun des droits spécifiques et des accès différents, ils sont représentés dans le schéma n°7 où la forme des bordures indique le type d'accès (libre, numéro IP de l'ordinateur ou un mot de passe).

Il y a plusieurs raisons à ce choix :

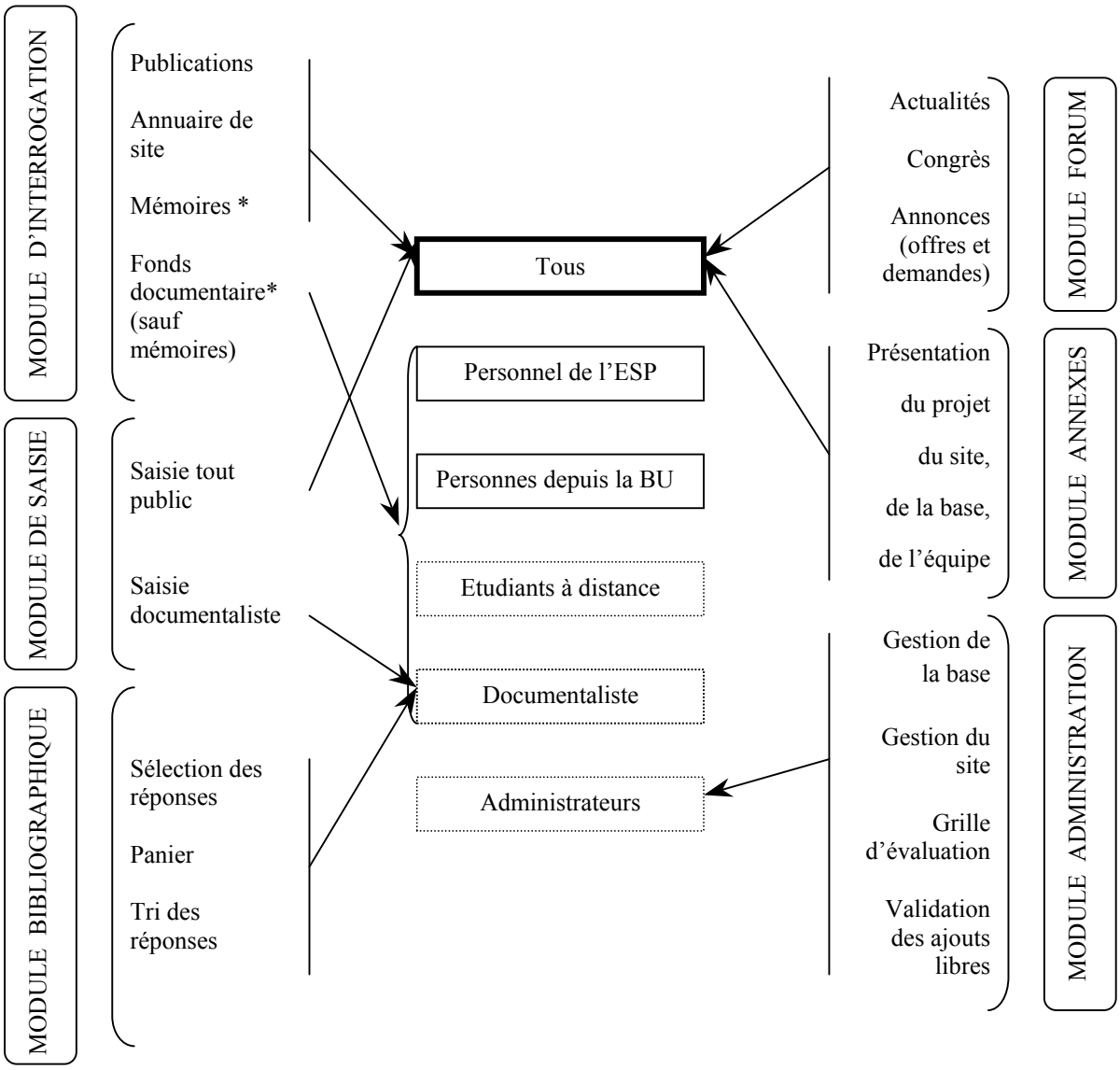
- ✧ Par définition, le module "Administration" n'est accessible qu'aux administrateurs du site.
- ✧ Comme la base SPIDER remplace l'ancien logiciel documentaire de l'école en propose les mêmes services, les fonctions de gestion du fonds restent uniquement accessibles par la documentaliste.
- ✧ Le catalogue du fonds est divisé en deux parties.

D'une part les références des mémoires internes sont en libre accès.

D'autre part les références des autres documents (livres, articles, congrès, CD) ne sont consultables que sur place par le personnel de l'école, les étudiants à la bibliothèque universitaire et à distance pour les étudiants en EAD (Enseignement à Distance) grâce à un accès protégé par un mot de passe.

Dans le cas de connexions extérieures à la faculté, les références fournies en réponses aux interrogations ont un contenu différent. En raison de la politique de prêt de documents de l'école qui souhaite satisfaire le maximum de demandes d'étudiants et pour des problèmes de disponibilités des documents, les références ne comportent pas d'indications sur la localisation dans le centre de documentation. A la place un guide est mis en ligne pour les aider à trouver, à proximité de chez eux, le document primaire correspondant à une référence.

SCHEMA n°7 : *LES DROITS ET LES ACCES DES DIFFERENTS TYPES UTILISATEURS*



L'astérisque * indique une différence dans le contenu des notices entre les connexions extérieures à la fac et les connexions intérieures, en effet la localisation de document n'est accessible que pour les personnes étant sur le campus.

Symbole des encadrés des types utilisateurs

Accès non réservé

Accès avec un mot de passe

Accès en fonction du numéro IP de l'ordinateur

3.3.2 Module d'interrogation

3.3.2.1 Formulaires d'interrogation

Pour créer les formulaires du site, j'ai étudié plusieurs formulaires de recherche provenant de sites en santé, de bases de données, de moteurs de recherche en français et en anglais. En comparant les champs interrogés, les modalités de recherche (case de texte, case à cocher, liste) et la présentation, plusieurs constations ressortent :

- Deux modalités d'interrogation sont souvent proposées ensembles :
  - un mode de **recherche simple** comprenant généralement une seule case de texte pour écrire une requête qui porte sur un ou plusieurs champs.
  - un mode de **recherche avancée** où l'interrogation se fait sur plusieurs critères.

- Deux types de recherche possibles :
  - en **texte intégral** quand la numérisation des documents est possible.
  - **recherche sur les champs** de la notice enregistrée dans la base.

- La présence d'une page d'aide à la recherche pour présenter les champs d'interrogation et les modalités d'écriture des requêtes : utilisation **d'opérateurs booléens** tel que ET, OU, SAUF, de **troncature**, la prise en compte ou non des **majuscules** et des **accents**, les règles de syntaxe du **parenthésage**...

- La présence de texte d'aide et de consignes de recherche sur le formulaire lui-même ou, parfois, des exemples de requêtes.

Les formulaires d'interrogation présents dans le site Santé-pub répondent à un certain nombre de contraintes :

- Ils doivent être rapides à remplir et d'utilisation intuitive ; un internaute doit pouvoir l'utiliser sans avoir besoin de lire toute l'aide avant. La page d'aide n'est utile que pour les néophytes ou pour pouvoir exploiter toutes les possibilités de recherche qui sont proposées.

- La saisie des requêtes doit être rapide. Dans les cas où le champ interrogé s'y prête, elle se fait grâce à une liste déroulante ou à l'aide d'un index, pour éviter les valeurs qui renverraient aucune réponse, et les fautes de frappes.

Un formulaire de recherche simple est proposé sur la page d'accueil (cf. Annexe 5.1) avec, à côté, un lien vers le formulaire de recherche avancée. La recherche avancée s'effectue à partir de deux formulaires.

Le formulaire de recherche sur l'annuaire de site ne propose que quelques champs d'interrogation (cf. Annexe 5.2).

Le formulaire de recherche pour les documents papier présente plusieurs modalités d'affichage variant selon l'utilisateur connecté.

✧ **un formulaire pour tout public** comprend tous les champs de recherches : choix de la partie de la base interrogée, choix du type de document, recherche libre multichamps, auteur, mots du titre, date de publication, mots-clés, diplôme (cf. Annexe 5.3).

✧ **un formulaire pour la documentaliste** se rajoute aux champs précédents des critères d'interrogation particuliers qui servent à la gestion du fonds : cote rangement, date de la fiche, numéro de la fiche, localisation, éditeur, collection, institution, catégorie, format d'affichage, critère de tri (cf. Annexe 5.4).

3.3.2.2 Page réponse et page d'aide

A côté des critères de recherche, il existe des paramètres de présentation de notice. Le formulaire d'interrogation en propose deux :

✧ Tri des réponses selon la date (ordre décroissant), le titre, l'auteur (ordre alphabétique), le type de document.

✧ Format d'affichage des références.

Les références bibliographiques s'affichent selon une présentation normalisée suivant la norme de Vancouver. Le texte de normalisation est issu de la "Convention de Vancouver : Recommandation aux auteurs pour des manuscrits devant être soumis à des revues biomédicales", organisé par le Comité International d'Organisation des Editeurs Médicaux et mis à jour en 1997 ; une large part de ce texte est consacré à l'écriture des références bibliographiques avec nombre d'exemples traités. Le choix s'est porté sur cette norme parmi toutes les autres existantes, car elle est employée à la fois par l'école (notamment pour la rédaction des mémoires d'étudiants) et par la faculté de médecine.

Quatre formats d'affichage ont été définis après concertation avec la documentaliste :

Format court : référence respectant à la lettre les règles (accessible par tous).

Format à distance : référence courte avec en plus les mots-clés et les zones géographiques (pays ou régions du monde traités dans le document). Ces champs supplémentaires aident la personne dans sa recherche. Ce format est donné par défaut pour les consultations à distance.

Format par défaut : même affichage que précédemment avec en plus la cote de rangement et la localisation du document. C'est le format donné par défaut pour les connexions à l'intérieur du campus de l'Ecole.

Notice éclatée : c'est le seul format qui présente la totalité des informations de la notice avec l'intitulé du champ et son contenu. Ce format correspond au format d'affichage pour vérifier la saisie d'une notice avant son enregistrement dans SPIDER.

Pour faciliter l'écriture de ces formats d'affichage, j'ai écrit un tableau indiquant pour chaque type de document les champs présents en fonction du format (cf. Annexe 6.2).

Pour la rédaction de notre page d'aide, chaque point traité comprend trois parties :

- ✧ un texte explicatif
- ✧ un exemple de requête
- ✧ le type de réponses obtenues

Le texte doit être explicite et court. Un internaute consacre, en général, peu de temps à la lecture des consignes. Parallèlement la navigation dans la page et entre la page et le formulaire est optimisée, ceci pour inciter les visiteurs à aller de l'un à l'autre. Ainsi chaque champ du formulaire de recherche avancée renvoie au niveau qui lui est consacré dans la page d'aide (cf. Annexe 5.5).

3.3.3 Module de saisie

Deux méthodes de saisie d'un document sont définies en fonction du type d'utilisateur comme indiqué dans le paragraphe.

3.3.3.1 Pour tout public

Les visiteurs du site ont à leur disposition quatre formulaires, un par type de documents : livres, congrès, articles, sites. Chacun regroupe l'ensemble des informations devant figurer dans les notices, une distinction est faite entre les champs obligatoires (s'ils ne sont pas remplis la saisie n'est pas enregistrée) et les champs facultatifs (cf annexes 7.1, 7.2, 7.3 et 7.4).

Les saisies enregistrées dans la base sont stockées pour être vérifiée et validée avant de devenir une notice interrogeable. L'étape de vérification et de validation des références est essentielle pour maintenir le niveau de qualité exigé par l'Ecole. Pour les livres, les articles, les congrès, la vérification peut se faire par comparaison avec des notices obtenues dans des bases de données tel que :

- ✧ **BDSP** (Bases de Données en Santé Publique) : base de données en santé publique, en libre consultation sur Internet, mais l'accès aux notices nécessite un abonnement,
- ✧ **MedLine** : base de données en Médecine accessible gratuitement depuis Internet,
- ✧ **Pascal** : base de données de l'INIST (Institut National d'Information Scientifique et Technique),
- ✧ Autres bases plus spécifiques comme **NOSOBASE** : Base de données sur l'hygiène hospitalière et la maîtrise des infections nosocomiales.

Ces sources permettent de vérifier l'existence du document et d'avoir sa notice mais les renseignements sur le contenu sont plus rares : MedLine donne cependant accès aux résumés des articles. L'indexation pose alors un problème, car si le centre de documentation ne possède pas le document, seul la personne proposant la référence peut l'effectuer.

Une solution à l'essai a été adoptée : proposés à la fois la saisie de mots libres et la saisie de mots-clés du thesaurus de la BDSP. Lors de la validation de la proposition, les mots libres sont traduits en concepts du thesaurus.

En effet on ne peut pas demander aux internautes d'indexer le document à partir du thesaurus complet (environ 3 000 termes), ni leur demander de ne donner que des termes au hasard.

L'alternative choisie est de faire un compromis : proposer trois exemples guidés avec une liste de mots tirés du thesaurus donnant une idée des grands thèmes traités et trois cases libres. Cette méthode d'indexation est celle adoptée par le site HON (Health On the Net Foundation) : fondation suisse à but non lucratif fournissant une base de données de sites médicaux "MedHunt" et une base de données de site web, d'articles scientifiques, d'actualité et de multimédia "HON select".

3.3.3.2. Pour la documentaliste

La documentaliste enregistre ses nouvelles acquisitions avec un formulaire unique pour tous ces documents : livres, congrès, articles, mémoires, CD-ROM, autres. Dès l'entrée dans la base, la notice devient interrogeable (cf. Annexes 7.5 e t 7.6).

Le recours à un formulaire unique laisse une grande souplesse d'enregistrement des références. Cette souplesse de saisie est une exigence de la documentaliste, car une grande partie de son fonds est composée de littérature grise[1].

Le catalogage de cette littérature est difficile car la présentation et les informations présentes ne sont pas standardisées, elles dépendent uniquement de l'organisme éditeur.

L'importation des données du logiciel documentaire GESBib dans SPIDER a mis en évidence quelques problèmes de saisie. Par exemple pour les sigles, plusieurs écritures sont possibles : avec les initiales ou avec le sigle développé où les majuscules ne sont pas toujours placées aux même endroits.

Le nouveau formulaire essaye de pallier ces divergences, il guide au mieux l'utilisateur avec des exemples et des consignes d'écriture pour chaque case et le recours à des listes déroulantes ou des index pour éviter les erreurs.

Les index utilisés sont inspirés par ceux présents dans le logiciel documentaire GESBib. Ils listent l'ensemble des valeurs d'un champ par ordre alphabétique dans une liste déroulante. Pour la parcourir, il suffit de taper les premières lettres du mot recherché.

[1] La littérature grise s'oppose à la littérature blanche qui représente les documents édités et imprimés en suivant des règles de présentation strictes. La littérature grise regroupe les documents à diffusion limitée, ne respectant pas de règles d'édition comme les mémoires, les thèses ou les rapports internes.

D'un point de vue technique, le langage HTML ne permet pas de réaliser ce type d'index. Une solution devra être adoptée pour pouvoir les concevoir, par exemple faire une interface avec ACCESS pour les formulaires ou bien définir une autre présentation, de la liste des valeurs d'un champ, rapide et efficace.

Comme pour le module de recherche, deux guides ont été rédigés un pour tout public, un pour la documentaliste où chaque champ renvoie à la partie du texte le concernant (cf. Annexe 7.7).

3.3.4. Module Bibliographie

Ce module est réservé à la documentaliste, il sert à écrire une bibliographie en réponse à la demande d'un membre du personnel de l'Ecole.

Le principe consiste à pouvoir sélectionner des références au sein de plusieurs pages de réponses obtenues après différentes recherches et de les stocker temporairement dans un panier. Là, elles sont triées selon plusieurs critères : date, titre, auteur, cote de rangement ou selon un ordre défini par la documentaliste (par un système de case de saisie où l'ordre est indiqué par un chiffre 1-2-…).

La liste ainsi établie peut être imprimée ou enregistrée dans un fichier texte. Ces fichiers peuvent être lus grâce à des logiciels de bureautique (comme Word de Microsoft) qui permettent des traitements de type tri et qui facilitent la mise en page des informations. La création de ces fichiers texte est intéressante pour regrouper des notices provenant de bases données différentes.

3.4 Evolution du site

Le site Santé-pub présenté ici est la première ébauche d'un projet qui peut aboutir à un site portail en santé publique. Pour atteindre cet objectif les services proposés doivent s'étoffer.

Plusieurs idées ont déjà été lancées :

- ✧ ajout un répertoire d'adresses de professionnels et d'organismes en santé publique,
- ✧ suivi des visites des sites de l'annuaire,
- ✧ installation d'un formulaire de vote qui informerait les administrateurs sur la qualité de la réponse fournie et sur le site visité,
- ✧ récupération des requêtes lancées pour mieux connaître les besoins des utilisateurs au niveau des documents et thèmes recherchés mais aussi sur les habitudes d'interrogation (champs et booléens utilisés).

Conclusion

L'objectif de ce stage était la mise en place de trois outils, la base données SPIDER divisée en trois parties : les publications, le fonds documentaire et l'annuaire de sites, la grille d'évaluation de sites en santé publique et le site web Santé-pub offrant, pour l'instant, un accès à SPIDER et à des forums de discussion.

En réponse à des contraintes techniques et humaines, ces outils doivent être simples à utiliser, faciles à gérer et évolutifs. Ils ont été conçus en réponses aux les attentes de l'Ecole de Santé Publique, et notamment de la documentaliste pour la partie fonds documentaire, et des besoins non formulés des professionnels et des étudiants en santé publique.

La principale incertitude concerne l'accueil du site par les internautes et plus particulièrement l'utilisation des forums de discussion, car nous n'avons aucun recul sur la question.

SPIDER et la grille en santé publique sont opérationnelles. Le contenue du site Santé-pub est défini plus ou moins précisément ; le module forum est en place, les modules interrogation, bibliographie et saisie sont prêts en version texte, pour les modules annexes et administration seuls quelques indications ont été données.

Au cours de ce stage, j'ai pu découvrir les multiples facettes de la participation à un projet : le travail en équipe, la définition des besoins de chacun, les réponses à y apporter en fonction de contraintes techniques, financières et humaines.

J'ai appris à planifier un travail que d'autres doivent concrétiser, pour la conception de la base de données SPIDER et du site en Santé Publique.

La partie du projet consacrée au site m'a permis d'exploiter l'outil Internet, car son utilisation est à la base des réflexions et des réalisations présentées ici : décryptage du contenu d'un site, des interfaces d'interrogation et de saisie conçues pour alimenter une base de données et la création d'une grille d'évaluation de sites web.

Parallèlement j'ai eu la chance de pouvoir participer au travail quotidien d'un centre de documentation et de me familiariser, un peu, avec les habitudes des utilisateurs, ce qui fut profitable pour mon travail.

La principale difficulté rencontrée consistait à appréhender les attentes et les besoins non formulés par les futurs utilisateurs du nouveau site et de leur apporter un outil de travail satisfaisant, aussi bien dans son contenu que dans sa forme.

Deux points se sont révélés plus épineux à traiter que les autres : premièrement, la prise de décision pour modéliser les fonctionnalités offertes par un logiciel documentaire, grâce à la mise en place de formulaire de saisie et d'interrogation et ceci en corrélation avec les besoins des internautes ; deuxièmement, le choix et surtout la formulation des questions de la grille d'évaluation qui est cruciale pour assurer la qualité de l'outil final.

L'Ecole de Santé Publique dispose de deux outils opérationnelles : la base SPIDER et la grille en santé publique.

A ce jour, le troisième outil Santé-pub est en cours de réalisation. La version "tout public" doit être mis en ligne au plus tôt. Les modules pour la documentaliste pourront être rajoutés ultérieurement.

L'avantage de ces outils est leur caractère évolutif qui leur permettra de suivre au mieux les besoins du public visé.

Ce projet est important pour le personnel de l'Ecole car il facilite l'accès à des informations qui, jusque là, n'étaient accessibles qu'à partir de quelques postes. Ainsi la mise en ligne offre de meilleures conditions de travail avec un gain de temps et des interfaces plus conviviales. Par ailleurs ce projet permet de mieux exploiter les compétences acquises par l'Ecole de Santé Publique.

Les services offerts, la base SPIDER et plus particulièrement l'annuaire de sites et le site Santé-pub, n'ont pas d'équivalent sur le web francophone. Les sites dédiés entièrement à la santé publique y sont rares dans le web francophone.

Par ailleurs Santé-pub a pour but non seulement de répondre aux besoins des professionnels en santé publique mais aussi des étudiants, cette double vocation me semble être le garant de l'avenir du projet.

Ce travail n'est que la première partie d'un projet plus vaste qui se déroulera sur plusieurs années. Le site Santé-pub modélisé ici ne représente que les fondations du projet. Les services offerts vont se diversifier : accès à un carnet d'adresses, mise en ligne d'un guide de recherche documentaire en santé publique. L'objectif est de le faire évoluer pour en faire un site portail[2] en santé publique.

[2] Site portail = site web dont la page d'accueil propose, en plus d'un moteur de recherche, des hyperliens avec une foule d'informations et de services attractifs, qui est conçu pour guider les internautes et faciliter leur accès au réseau, mais surtout pour les attirer et fidéliser le plus grand nombre d'entre eux, au point de devenir leur porte d'entrée dans l'Internet. (http://www.olf.gouv.dc.ca/ressources/internet/fiches/8390683.html)

BIBLIOGRAPHIE[3]

Marie Françoise de PANGE. Internet : la bataille des portails de santé. Le Quotidien du Médecin 14 décembre 1999 : 3-5.

Tabozzi M, Osèle S. Internet et Santé Publique : accès à l'information. Santé publique 1999 ; 11 (3) : 363-367.

Thirion B, Darmoni S. L'Internet, les bibliothèque et la médecine : l'expérience de CHU de Rouen. Documentaliste – Sciences de l'information 1996 ; 33 (3) : 171-175.

Tabozzi M. Formation des médecins de l'Education Nationale : Ressources en santé sur Internet. Document semaine 1999 ; (3) : 46.

Support de cours : Introduction à la santé publique et communautaire, dossier documentaire. Nancy : Ecole de Santé Publique ; 2000. 213p.

Banque de données en Santé Publique. Thesaurus : liste alphabétique permuté version2.0. EDITIONS ENSP ; 1995. 256 p.

Banque de données en Santé Publique. Thesaurus : liste alphabétique structurée version2.0. EDITIONS ENSP ; 1995. 256 p.

Banque de données en Santé Publique. Thesaurus : terminogrammes et listes alphabétiques par microthesaurus version2.0. EDITIONS ENSP ; 1995. 256 p.

Charton E. Site Internet : Conception et réalisation. Editeur : S&SM ; 1997. 317 p.

Herellier JM, Wolosin C. Bien référencer son site Internet. Editeur : S&SM ; 1998. 226 p.

Leloup C. Moteurs d'Indexation et de Recherche : Environnement client - serveur, Intranet et Internet. Eyrolles ; 1998. 281 p.

Cottet P. Réussir son site Web. Paris : Osman Eyrolles Multimédia ; 1999. 314 p.

Gray D. Optimisez vos pages Web. Paris : Osman Eyrolles Multimédia ; 2000. 260 p.

Institut National de Recherche en Informatique et en Automatique. Créer et maintenir un site Web : cours de l'INRIA du 28 septembre – 2 octobre 1998 à Pau. Paris : ADBS Editions : 1998. 294 p.

Clay A. Architecture Web. Paris : Edition First Interactive ; 2000. 216 p.

[3] Les références sont présentées selon les normes de Vancouver. Elles sont présentées dans l'annexe 6.1.

LISTES DES SITES VISITES[4]

Les sites portails :

HON Health On the Net Foundation, [en ligne]. http://www.hon.ch/ (page consultée le 28 août 2000)

Promédica. L'Internet de la santé, [en ligne]. http://www.promedical.net/ClubPro/pagesSites/index.html

(page consultée le 28 août 2000)

Caducee. La santé sur Internet, [en ligne]. http://www.caducee.net/ (page consultée le 28 août 2000)

Atmedica. Tout l'Internet médical, [en ligne]. http://www.atmedica.com/

Medisite. L'Internet au service de la médecine, [en ligne]. http://www.medisite.fr/ (page consultée le 28 août 2000)

CISMef. Catalogue et index des sites médicaux francophones, [en ligne]. http://www.chu-rouen.fr/cismef/ (page consultée le 28 août 2000)

WebMed. Web guide médical, [en ligne, avec mot de passe]. http://www.webmedguide.com/ (page consultée le 28 août 2000)

Egora. La santé en direct, [en ligne]. http://www.egora-sante.com/NewAccueil.asp (page consultée le 28 août 2000)

Visulogic. Les nouveaux portails médicaux, [en ligne]. http://www.visulogic.com/sommaire.htm

IndexMed. Un site du groupe Impact médecin, [en ligne]. http://www.indexmed.com/ (page consultée le 28 août 2000)

Med in Web. La santé analysée par les experts, [en ligne]. http://www.medinweb.fr/web/home.php

(page consultée le 28 août 2000)

Les moteurs de recherches :

Yahoo. Yahoo France, [en ligne]. http://fr.yahoo.com/ (page consultée le 28 août 2000)

Ecila, [en ligne]. http://query.ecila.fr/ (page consultée le 28 août 2000)

Google. Google français, [en ligne]. http://www.google.com/intl/fr/(page consultée le 28 août 2000)

Hotbot, [en ligne]. http://hotbot.lycos.com/ (page consultée le 28 août 2000)

[4] La présentation de ces références est faite avec les conseils données par l'Université de LAVAL
Caron, Rosaire. "Comment citer un document électronique ?". In Université Laval. Bibliothèque. Site de la Bibliothèque de l'Université Laval, [en ligne]. http://bibl.ulaval.ca/doelec/citedoce.html (Page consulté le 26 juin 2000)

Go.com, [en ligne]. http://www.go.com/ (page consultée le 28 août 2000)

Altavista. Altavista recherche France, [en ligne]. http://fr.altavista.com/ (page consultée le 28 août 2000)

Altavista. Altavista search Canada, [en ligne]. http://www.altavistacanada.com/cgi-bin/query (page consultée le 28 août 2000)

WebCrawler. WebCrawler It's that simple, [en ligne]. http://www.webcrawler.com/ (page consultée le 28 août 2000)

Excite. Excite turn you on, [en ligne]. http://www.excite.com/ (page consultée le 28 août 2000)

Les bases de données et catalogues :

NCBI. National Library of Medecine Pub Med, [en ligne]. http://www.ncbi.nlm.nih.gov/PubMed/ (page consultée le 28 août 2000)

National Library of Medecine. Grateful Med, [en ligne]. http://igm.nlm.nih.gov/ (page consultée le 28 août 2000)

ScienceDirect. Welcome to ScienceDirect – your Web databse for scientific, tecnical and medical research, [en ligne]. http://www.sciencedirect.com/science/home (page consultée le 28 août 2000)

BDSP Banque de Données en Santé Publique, [en ligne]. http://www.bdsp.tm.fr/ (page consultée le 28 août 2000)

NOSOBASE Base de données sur l'hygiène hospitalière et la maîtrise des infections nosocomiales, [en ligne]. http://www.univ-lyon1.fr/LyonSud/nosobase/index.htm (page consultée le 28 août 2000)

Université Henri Poincaré. Service Commun de Documentation, [ligne]. http://scd.uhp-nancy.fr/SCD/catal.htm (page consultée le 28 août 2000)

Université de Nancy II. Service Commun de Documentation, [en ligne]. http://www.univ-nancy2.fr/webbib/ (page consultée le 28 août 200)

Les sites en santé :

Santécom. Service documentaire québécois spécialisé en santé publique, [en ligne]. http://www.santecom.qc.ca/ (page consultée le 28 août 2000)

Sentiweb, [en ligne]. http://www.b3e.jussieu.fr:80/sentiweb/ (page consultée le 28 août 2000)

ENSP Ecole National en Santé Publique, [en ligne]. http://www.ensp.fr/ (page consultée le 28 août 2000)

Santé Canada Santé Publique, [en ligne]. http://www.hc-sc.gc.ca/francais/sante.htm (page consultée le 28 août 2000)

Santé Canada. Direction générale de la santé de la population et de la santé publique (anciennement LLCM), [en ligne]. http://www.hc-sc.gc.ca/hpb/lcdc/public_f.html#top (page consultée le 28 août 2000)

Institut de Veille Sanitaire, [en ligne]. http://www.rnsp-sante.fr/ (page consultée le 28 août 2000)

WHO/OMS. World Health Organisation /Organisation Mondiale De La Santé, [en ligne]. http://www.who.int/ (page consultée le 28 août 2000)

Les sites pour webmasters :

WebDeveloppeur.com. Le guide des webmasters francophones, [en ligne]. http://www.webdeveloppeur.com/ (page consultée le 28 août 2000)

Commission nationale de l'informatique et des libertés, [en ligne]. http://www.cnil.fr/ (page consultée le 28 août 2000)

Comment ça marche ?[l'informatique], [en ligne]. http://www.commentcamarche.net/ (page consultée le 28 août 2000)

UNGI Un Nouveau Guide Internet, [en ligne]. http://www.imaginet.fr/ime/fr_ungi2.htm (page consultée le 28 août 2000)