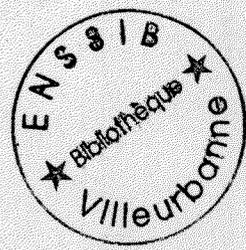


ECOLE NATIONALE SUPÉRIEURE
DES SCIENCES DE L'INFORMATION
ET DES BIBLIOTHEQUES

UNIVERSITÉ
CLAUDE BERNARD
LYON I

D.E.S.S. en INFORMATIQUE DOCUMENTAIRE

Rapport de stage

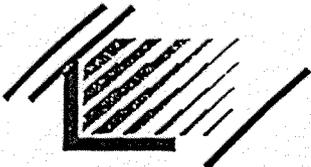


ELABORATION D'UN SERVEUR WEB ET D'UNE BASE WAIS
A LA BIBLIOTHEQUE DU CNET DE GRENOBLE

Valérie TRILLAT
Sous la direction de
Christine TRAYNARD

Centre National d'Etudes des Télécommunications de Grenoble
28 chemin du Vieux Chêne - BP 98
38 243 Meylan Cedex
- 1995 -

1995
ID ST
18



e n s s i b

école nationale supérieure des sciences de l'information et des bibliothèques

Afin de ne pas pénaliser nos lecteurs,
veuillez s'il vous plait respecter les délais de prêt.

Prêt jusqu'au :

5 / 12

ECOLE NATIONALE SUPÉRIEURE

UNIVERSITÉ

DES SCIENCES DE L'INFORMATION

CLAUDE BERNARD

ET DES BIBLIOTHEQUES

LYON I

D.E.S.S. en INFORMATIQUE DOCUMENTAIRE

Rapport de stage



ELABORATION D'UN SERVEUR WEB ET D'UNE BASE WAIS
A LA BIBLIOTHEQUE DU CNET DE GRENOBLE

Valérie TRILLAT
Sous la direction de
Christine TRAYNARD

Centre National d'Etudes des Télécommunications de Grenoble
28 chemin du Vieux Chêne - BP 98
38 243 Meylan Cedex
- 1995 -

1995

ED ST

18

Sujet : Elaboration d'un serveur Web et d'une base Wais à la bibliothèque du CNET de Grenoble.

Valérie TRILLAT

Résumé : Ce rapport de stage décrit les étapes d'élaboration et les outils employés pour mettre en place un serveur World Wide Web et une base Wais contenant des notices bibliographiques à la bibliothèque technique du Centre National d'Etudes des Télécommunications de Grenoble.

Descripteurs* : hypertexte, hypermedia, serveur, Internet, World Wide Web, W3, WWW, Web, Wais, langage HTML, catalogue.

Abstract : This report describes the development and tools used to create a World Wide Web server and a Wais database for the technical library of the Centre National des Télécommunications (National Research Center for Telecommunications) of Grenoble.

Keywords : hypertext, hypermedia, server, Internet, World Wide Web, W3, WWW, Web, Wais, HTML language, catalogue.

* seuls les descripteurs soulignés sont issus du thésaurus Pascal, les autres appartiennent au langage naturel

Remerciements

Je souhaite remercier ici Christine Traynard, qui m'a accueillie à la bibliothèque et a toujours su, en dépit d'un éloignement forcé, prêter un oeil attentif à mon travail.

Je remercie sincèrement Philippe Maurice pour son aide et ses conseils avisés.

Je remercie vivement Jean-René Brosselin pour son précieux concours et sa patience.

Merci à Roland Gerber pour m'avoir fait confiance.

Merci à Sylvie Jimenez pour sa gentillesse.

Un clin d'oeil à mes collègues stagiaires et intérimaires Sean, Véronique, Julie et Frédéric pour l'ambiance agréable et le café du matin.

Une dernière pensée pour Geneviève en souvenir de nos fructueux échanges téléphoniques.

Enfin, merci à Elisabeth Cherchal, documentaliste au CICG.

TABLE DES MATIERES

Introduction	6
I . Rappels sur internet	8
1 . Historique	9
2 . Evolution récente	9
3 . Perspectives futures : Le W3C	12
II . Etude préliminaire et organisation des contenus	14
1 . Définition du projet	15
2 . Prospection sur le Web	15
3 . Choix des contenus	16
3.1 Informations générales	16
3.2 Services proposés	16
3.2.1 Informations sur les abonnements	16
3.2.2 Listes des acquisitions récentes	16
3.2.3 Fourniture de documents	16
3.2.4 Accès aux bases de données	17
3.2.5 Accès à la documentation interne	17
a. Formulaire simple	17
b. Formulaire détaillé	17
c. Contenu des notices	18
3.3 Ressources extérieures	18
III. Le serveur World Wide Web	19
1. Présentation générale	20
1.1 Schéma du serveur du CNS	21
1.2 Comment fonctionne un serveur web?	22
1.2.1 L'architecture client/serveur	22
1.2.2 Le protocole HTTP	23
a. Fonctionnement d'HTTP	23
b. L'adressage HTTP	23
1.2.3 Les URLs	23
a. URL relatif	24
b. URL absolu	24
1.3 Le logiciel serveur	24
1.4 Le logiciel client	24
2. Réalisation du serveur	25
2.1 Les documents HTML	25
2.1.1 La norme SGML	25
2.1.2 La DTD HTML	25
2.1.3 Les balises HTML	26
a. Les En-têtes	26
b. Les délimiteurs	26
c. Les styles	26
d. Caractères accentués et symboles	27
e. Les listes	27
f. Les images dans le texte	27

g. Les images réactives	28
h. L'attribut background	29
i. Les liens hypertextes	29
j. Appel d'un programme	30
2.1.4 Quelques règles d'ergonomie	37
Organnigramme du serveur	38
2.2 Les outils de rédaction	39
2.2.1 Les éditeurs HTML	39
2.2.2 Les gabarits HTML pour Word	39
2.2.3 Les convertisseurs	39
2.3 L'insertion des images fixes	40
IV . La base Wais	42
1 . Présentation générale	43
1.1 Comment fonctionne une base Wais?	43
1.1.1 L'architecture client/serveur	43
1.1.2 Le protocole Z39.50	43
1.2 Le logiciel d'indexation : waisindex	44
1.3 Le logiciel serveur : waisserver	44
1.4 Le logiciel client : waisq	45
1.5 Le niveau de diffusion de la base	45
2 . Réalisation de la base	45
2.1 Objectifs de la base	45
2.2 Stockage de la base	46
2.3 Mode d'interrogation	46
2.3 Problèmes d'intégration de documents complets	46
2.4 Passerelle Wais/Web	47
V . Maintenance et promotion	48
1 Règles rédactionnelles pour les pages html	49
2 Organisation des fichiers	49
3 Aide à la mise à jour de la base Wais	50
3.1 Consignes de rédaction des notes techniques	50
3.2 Programmes pour saisir des modifications	50
3.2.1 Remnotice	50
3.2.2 Addnotice et modnotice	50
4 . Faire connaître le serveur	51
Conclusion	53
Annexes	54
Bibliographie	63

INTRODUCTION

Présentation du CNET

France Télécom, quatrième opérateur mondial en télécommunications, gère un réseau souvent présenté comme l'un des plus performants du monde. C'est grâce à son centre de recherches, le CNET (Centre National d'Etudes des Télécommunications), que France Télécom a pu se hisser à ce niveau de compétences.

Créé en 1944, le CNET joue un rôle moteur dans la modernisation des télécommunications françaises et contribue au développement de ces industries.

Avec un budget d'environ 2,5 milliards de francs et fort de 4000 chercheurs répartis sur six centres, le CNET est à l'avant garde du groupe chargé de la prospection technologique.

Ouvert en 1978, le CNET de Grenoble (aussi appelé CNS, Centre Norbert Segard) a pour mission la recherche en micro-électronique sur le silicium. Installé depuis 1981 sur la ZIRST de Meylan, technopole de la région grenobloise, le CNS emploie environ deux cents chercheurs qui ont à leur disposition 8000 M² de laboratoire.

Depuis juillet 1990, France Télécom a un statut d'exploitant public autonome. Il est à la fois opérateur libre sur un marché concurrentiel et entreprise protégée responsable du développement d'un service public. Toutefois, la Commission Européenne a décidé qu'à partir de 1998, le marché des services publics serait ouvert à la concurrence.

Dans ce contexte, le CNET doit assurer une veille technologique efficace qui passe par le développement de ses centres de documentation.

La bibliothèque technique du CNS

Directement rattachée à la direction du CNS, la bibliothèque technique tient donc une place prépondérante dans la vie du centre.

Son fonds rassemble 7500 ouvrages et 280 titres de périodiques. Par ailleurs, celle-ci fait partie d'un réseau documentaire qui met en commun les fonds des huit bibliothèques des centres de recherches et Ecoles des Télécommunications. Ce fonds est géré par le logiciel Dobitel, qui contient les références de 150 000 monographies. Un nouveau logiciel, doit être mis en place courant 1996.

Début 1995, un stagiaire de l'IUT Informatique de Grenoble a développé un serveur Web pour le CNS. Il a pour principal objectif d'améliorer la communication interne en même temps qu'il doit permettre aux chercheurs de découvrir l'Internet.

Concernant la bibliothèque, les pages du serveur devront permettre à l'utilisateur d'obtenir divers renseignements (modalités de fonctionnement, sources disponibles...) et d'effectuer différentes opérations depuis son terminal (commande de photocopies, demande de PEB, interrogation d'une base Wais contenant les notes techniques ...).

Précision importante, le serveur du CNET n'est actuellement pas accessible de l'extérieur, il est réservé à un usage interne, en attendant une éventuelle autorisation d'ouverture.

I. RAPPELS SUR INTERNET

1. HISTORIQUE

En 1969, le ministère de la défense des Etats-Unis lance un projet de développement et de recherche pour créer le réseau Arpanet. Ce réseau devait être fiable et indépendant des constructeurs. Devant son succès, les organismes associés à son développement l'utilisent pour la transmission de leurs données. En 1975, Arpanet devient opérationnel sous la direction de l'Agence de Communication du ministère de la Défense.

En 1983, le protocole TCP/IP¹ est adopté comme standard militaire. La même année, le terme d'Internet fait son apparition et s'appuie sur le protocole TCP/IP.

Aujourd'hui, Internet est un réseau de réseaux connectés entre eux de façon anarchique. Ces réseaux peuvent être nationaux, régionaux ou limités à une institution.

75 pays sont reliés par Internet ce qui représente près de 3 millions de serveurs et 40 millions d'utilisateurs. Sa croissance annuelle est d'environ 10% par an.

En Europe, Ebone est le principal réseau fédérateur IP (Internet Protocol). Son architecture est organisée en quelques grands noeuds de communication dont Paris, le CERN, Amsterdam, Stockholm et Vienne. Les liaisons entre ces noeuds sont à haut débit : 1 ou 2Mbits/s. Il y a aussi des liaisons à haut débit avec les Etats-Unis, elles permettent de communiquer avec l'Australie, le Japon et l'Asie du sud-est. Le noeud parisien de Ebone est hébergé chez France Télécom.

2. EVOLUTION RÉCENTE

Jusque très récemment, les services disponibles sur Internet se résumaient aux services de base. Des serveurs ftp anonymes contenaient logiciels et documentation, certaines machines offraient une entrée en "telnet anonyme" pour permettre l'interrogation de bases de données.

Mais depuis deux ans environ sont apparus de nouveaux services : Wais, Gopher, et surtout World Wide Web qui soulève une vague d'enthousiasme et poursuit une croissance exponentielle. Cette popularité résulte des possibilités nouvelles offertes par ce système d'information qui vise un public très large, bien au-delà du monde des informaticiens.

Ces serveurs, mis en place de façon décentralisée, par la base, souvent dans le but de rendre service à une communauté locale, ne possèdent aucune organisation commune permettant de retrouver ce

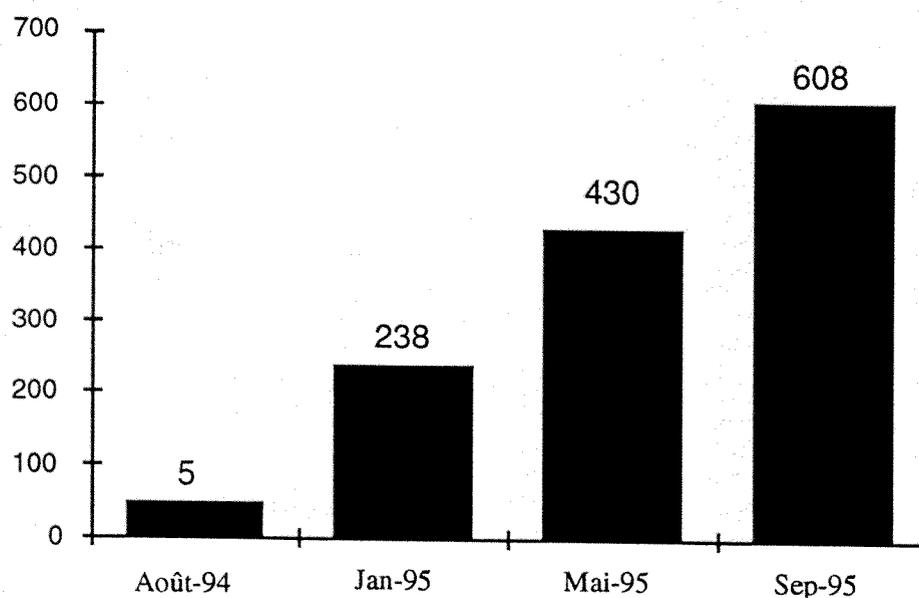
¹ Transmission Control Protocol / Internet Protocol

Série de protocoles d'interconnexion de réseau spécifiant un niveau de transport et un niveau de routage. Il est destiné au support d'applications de type transactionnel ou au transport de fichiers

que l'on cherche. La jeunesse de ces services et la forte croissance du nombre de machines connectées donnent une impression de "pagaille" dans laquelle il est difficile de trouver ce que l'on cherche.

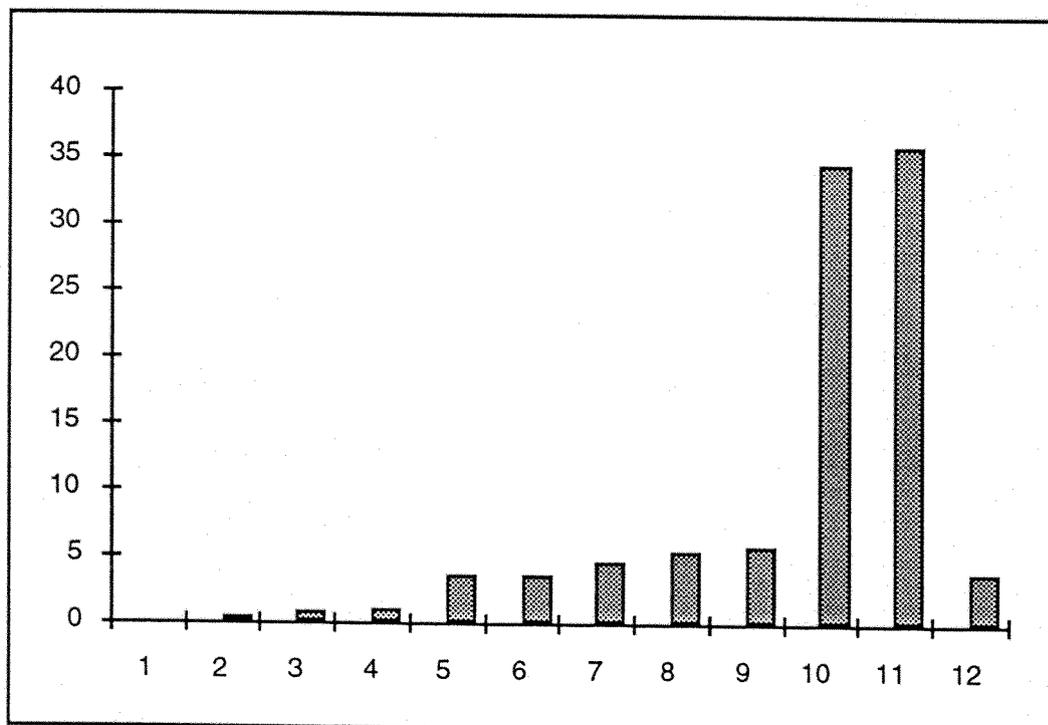
C'est pourquoi de nombreux projets de recherche travaillent actuellement à une organisation de l'information de l'Internet dans le but de faciliter la recherche par l'utilisateur.

EVOLUTION DU NOMBRE DE SERVEURS WEB EN FRANCE



(d'après des chiffres du serveur du CNRS disponibles à l'adresse suivante <http://www.urec.fr>)

RÉPARTITION DES SERVEURS WEB
PAR SECTEURS D'ACTIVITÉS



- 1 : Lycées : 0,024%
- 2 : Organismes internationaux : 0,48%
- 3 : Presse : 0,73%
- 4 : Administration centrales (hors culture) : 0,97%
- 5 : Collectivités locales, territoriales : 3,63%
- 6 : Secteur culturel : 3,63%
- 7 : Prestataires Internet : 4,6%
- 8 : Entreprises : 5,33%
- 9 : Associations : 5,81%
- 10 : Recherche : 34,62%
- 11 : Universités : 36,08%
- 12 : Autres : 3,87%

(D'après Solutions Télématiques n°37, juillet/août 1995)

3. PERSPECTIVES FUTURES : LE W3C²

WORLD WIDE WEB CONSORTIUM

Le consortium international pour le Web, fondé en octobre 1994 par le Massachusetts Institute of Technology (MIT) est dirigé par Tim Berners-Lee, inventeur du Web au CERN en 1993.

Jusqu'au début de 1995, le CERN dirigeait le W3C aux côtés du MIT puis il fut remplacé par l'INRIA qui en est désormais le pôle européen.

Un troisième partenaire, asiatique, devrait bientôt rejoindre la direction.

Cette organisation, dont la durée de vie est pour le moment limitée à 1997, compte une soixantaine de membres qui la financent par des droits d'adhésion. Parmi eux, des opérateurs de télécommunications (ATT, Bellcore, NCSA...), des éditeurs de logiciels (Microsoft, Lotus, Novell, Netscape...), des constructeurs informatiques (IBM, Sun, Fujitsu...) et des fournisseurs de contenus (CompuServe, O'Reilly...).

Le rôle de cette organisation est de définir les standards du Web (URL, HTTP, HTML notions définies plus loin), de concevoir les protocoles, d'encourager les entreprises à créer des produits conformes à ces protocoles et de promouvoir ces produits.

Le W3C coordonne notamment les travaux sur HTML+, langage de rédaction des documents du Web, qui est en cours de finalisation. L'évolution vers un Web tridimensionnel relève également de ses compétences.

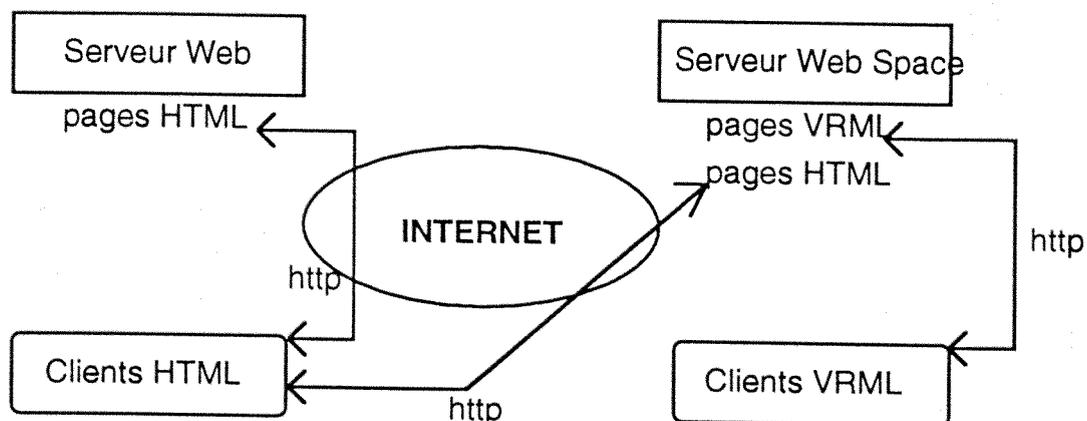
Actuellement, le Web gère des documents en 2D mais une initiative a été lancée par Silicon Graphics et Templates Graphics Software (TGS). Baptisé Web Space, ce concept est fondé sur le langage VRML (Virtual Reality Markup Language), défini par Templates Graphics Software sur la base d'Open GL, une interface de programmation d'applications 3D créée par Silicon Graphics.

D'autres développements ajoutent l'interaction entre multi-utilisateurs dans un interface 3D. Ils permettent aux utilisateurs de communiquer et d'agir dans un espace où ils sont représentés par leur "image virtuelle".

A terme, cette technologie permettra de fournir la présence virtuelle d'une entreprise sur Internet.

Cette évolution doit toutefois tenir compte de plusieurs restrictions, il faudra notamment être connecté par un ordinateur très puissant et disposer d'une liaison à haut débit pour accéder correctement à un environnement en 3D.

² le serveur du World Wide Web Consortium est accessible à l'adresse suivante:
<http://www.w3.org/#W3C>

LES LIAISONS DE WEB ET DE WEB SPACE

Ces deux types de réseaux sont liés car le protocole qui permet les échanges entre le client et le serveur est HTTP dans les deux cas. D'autre part, un serveur Web Space est capable de traduire des pages VRML en pages HTML. Une compatibilité ascendante avec le Web classique est ainsi assurée.

II. ETUDE PRELIMINAIRE ET ORGANISATION DES CONTENUS

1. DEFINITION DU PROJET

Le serveur de la bibliothèque doit s'intégrer au serveur du CNET réalisé par un stagiaire d'IUT d'informatique au début de l'année 1995.

Les pages concernant la bibliothèque (que nous appellerons désormais le serveur de la bibliothèque) seront accessibles depuis le menu principal du serveur du CNS, par un icône.

Ce serveur doit être un outil efficace pour les chercheurs, leur apportant un maximum d'informations et de documentation directement sur leur poste.

La documentation disponible en ligne sera rassemblée dans une base Wais. Dans un premier temps, seules les notes techniques (rapports des chercheurs et rapports des stagiaires) seront proposées dans cette base. Plus tard, les thèses et publications de chercheurs y seront intégrées.

Cette base sera interrogée depuis le serveur Web par l'intermédiaire d'une passerelle Web/Wais.

Le serveur doit également être une ouverture sur les ressources documentaires disponibles sur Internet.

2. PROSPECTION SUR LE WEB

Les Ecoles Nationales Supérieures des Télécommunications de Paris et de Bretagne (ENST et ENSTB), ainsi que les autres CNET ont développés leurs propres serveurs Web au sein desquels une place est faite aux bibliothèques.

Nous nous sommes donc intéressés au mode de présentation de l'information sur chaque site ; sachant notamment que les autres CNET traitent le même type de documents et participent au réseau documentaire commun Dobitel.

Les serveurs d'autres centres de recherche réservent également une large place à leurs bibliothèques, nous les avons donc consultés afin de définir les règles d'organisation de nos propres pages.

Quelques uns des serveurs qui nous ont inspirés sont :

celui de l'Ecole Nationale Supérieure des Télécommunications de Paris (<http://web.enst.fr/>),

de l'Ecole Nationale Supérieure des Télécommunications de Bretagne (<http://www.enst-bretagne.fr/>),

de REDOC (Réseau Documentaire grenoblois) (<http://www.grenet.fr>)

du Centre Interuniversitaire de Calcul de Grenoble (<http://www.grenet.fr/cicg/cicg.html>),

de l'IMAG (Information Mathématiques Appliquées de Grenoble) (<http://www.com.grenet.fr:80/anteserveur/imag...>)

3. CHOIX DES CONTENUS

De ces pérégrinations sur le Web est ressortie l'observation suivante : lorsque des pages sont réservées à la bibliothèque, l'information y est fréquemment présentée selon trois axes.

Tout d'abord les renseignements pratiques, utiles notamment aux nouveaux arrivants sur le serveur ; puis le détails des services disponibles et enfin l'accès à des ressources extérieures, sur d'autres serveurs Web de l'Internet.

Par ailleurs, de plus en plus de serveurs de bibliothèques proposent leur documentation en ligne stockée dans une base Wais contenant les notices et parfois les documents complets.

Enfin, pour choisir les informations à intégrer, nous disposions également, comme base de travail, du guide papier réalisé en 1993 par les documentalistes.

3.1...INFORMATIONS GÉNÉRALES

Nous avons donc suivi ce schéma de présentation. Le serveur donne tout d'abord des informations très générales, qui seront utiles aux thésards, jeunes ingénieurs, stagiaires...

Depuis la première de ces pages, l'utilisateur peut accéder à un guide rassemblant les principales informations et imprimable en un seul document.

3.2...SERVICES PROPOSÉS

Il a fallu ensuite définir quels seraient les services proposés aux utilisateurs afin de répondre au mieux aux exigences de la direction (fournir aux chercheurs le maximum de renseignements et de documents en ligne depuis leurs postes) tout en facilitant le travail des documentalistes grâce à l'automatisation de certaines procédures.

Ces services peuvent être divisés en cinq groupes :

3.2.1 informations sur les abonnements

Une liste alphabétique et une liste thématique des abonnements sont proposées. La liste thématique présente dix ancres, une par groupe de la classification décimale. De plus, l'utilisateur dispose d'une grille d'interrogation afin de savoir si la bibliothèque est ou a été abonnée au titre qui l'intéresse.

3.2.2 listes des acquisitions récentes

Il s'agit également de listes thématique et alphabétique qui seront mises à jour chaque mois.

3.2.3 fourniture de documents

Derrière cette rubrique se trouvent trois formulaires à remplir par l'utilisateur : le premier pour les demandes de photocopies d'articles, le second pour les prêts du réseau Dobitel ou PEB, le dernier pour les suggestions d'achat.

Jusqu'ici, la demande écrite du lecteur était recopiée pour être ensuite envoyée aux fournisseurs.

Notre outil permettra à l'utilisateur de remplir sa demande depuis son bureau.

Dans un premier temps, les renseignements seront intégrés, si possible à un tableau Excel, ou tout au moins envoyé automatiquement par courrier électronique une fois par jour.

Par ailleurs il serait utile de développer un outil permettant de récupérer les renseignements dans un tableau dont le format soit conforme aux formulaires des fournisseurs (formulaire de l'INIST par exemple), ce qui représenterait un gain de temps appréciable.

3.2.4 accès aux bases de données

Cette partie du serveur présente différentes sources sur support électronique disponibles à la bibliothèque (Bases en ligne, CD-ROM, Current Contents) et décrit leurs modes d'interrogation. Elle détaille également le type de renseignements que l'on peut y trouver.

3.2.5 accès à la documentation interne

Il s'agit de la base Wais pour laquelle deux formulaires d'interrogation sont prévus.

a. formulaire simple

Le premier, assez simple, doit permettre au nouvel usager de se familiariser à l'interrogation (se reporter à l'annexe I). Il ne propose que trois types de champs : *auteur*, *terme* (qui peut être un mot du titre ou un mot du résumé) et *site de publication*.

L'utilisateur sait parfaitement ce que renferment ces types de champs. Par ailleurs, ceux-ci sont reliés par l'opérateur booléen ET. Pour les champs *auteur* et *terme*, trois cases sont proposées. Entre des cases de même type il y a un OU afin de ne pas trop restreindre l'interrogation et permettre d'interroger sur plusieurs auteurs ou plusieurs termes.

b. formulaire détaillé

Le second formulaire d'interrogation est plus complexe, vous pouvez le consulter en annexe II. Il propose d'interroger sur neuf champs différents :

- auteur (trois cases)
- mot du titre
- descripteurs (trois cases)
- année
- type de document
- nom du responsable (pour les rapports de stage)
- mots du résumé (trois cases)
- site de publication
- groupement

Un ET relie toutes les cases entre elles. Les cases *type de document* et *site de publication* présentent des menus déroulants dans lesquels

l'utilisateur clique sur l'option de son choix. L'option "non connu" permet de laisser ces champs à vide.

Ce formulaire n'est pas fait pour être rempli intégralement mais plutôt pour permettre des combinaisons très variées d'interrogation. En effet, une interrogation sur tous les champs n'apporterait vraisemblablement jamais de réponse, étant donné le volume réduit de la base.

Pour les deux formulaires, il est nécessaire de donner quelques explications avant la première utilisation, notamment pour interdire l'usage des caractères accentués, préciser ce qu'est un descripteur, que les chiffres ne peuvent être saisis seuls etc...

c. contenu des notices

La base Wais étant composée des notices bibliographiques de notes techniques, il faut décider quelles informations doivent figurer dans une notice.

Sachant que les notes techniques sont toujours le fruit du travail de plusieurs chercheurs qui appartiennent parfois à des groupements³, voire à des sites différents, certaines informations s'avèrent indispensables.

Hormis les noms d'auteurs, le titre du document, l'année de rédaction, le résumé et les descripteurs, il faut donc préciser le(s) groupement(s), le type du document (note technique ou rapport de stage), le(s) site(s) de publication et le numéro attribué par l'administration qui fera office de numéro de référence. Il sera utilisé pour nommer le fichier contenant le texte intégral de la note technique.

Pour les rapports de stage, il faut ajouter un champ nom du responsable qui permettra plus aisément de retrouver un rapport quand le nom du stagiaire-auteur aura été oublié.

3.3 RESSOURCES EXTERIEURES

Dans cette troisième partie, des liens permettent d'accéder directement à des serveurs extérieurs. Ils traitent de télécommunication et de microélectronique.

Un lien propose également d'accéder au serveur *Le Monde en Tique* tenu par la librairie du même nom, spécialisée en informatique.

Un autre serveur est également pointé, celui du Réseau Documentaire Grenoblois (REDOC) dont fait parti la bibliothèque du CNET. Il propose notamment un annuaire des centre de documentation de la région.

Enfin, la bibliothèque étant le lieu de stockage des cartes et du matériel de randonnée, il m'a semblé intéressant de créer un lien en direction de CIMES, le serveur sur les Alpes développé par ma collègue Geneviève Escomel.

³ chaque CNET est divisé en plusieurs groupements. Par exemple la bibliothèque appartient au groupement PCV (Promotion, Communication, Valorisation)

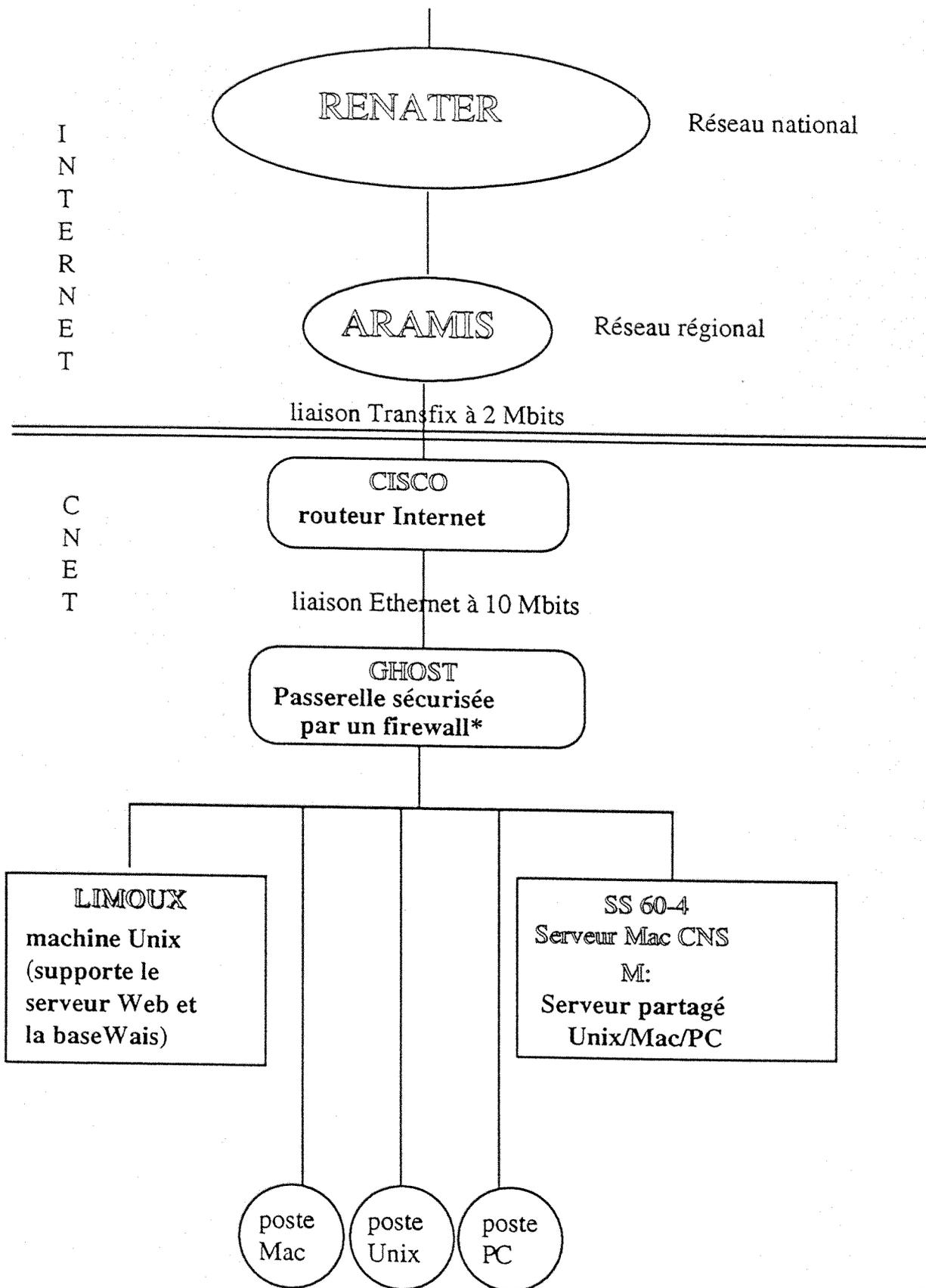
III. LE SERVEUR WORLD WIDE WEB

1. PRÉSENTATION GÉNÉRALE

Le projet Web a été développé en 1993 au CERN par Tim Berners Lee. Il est basé sur le protocole HTTP (HyperText Transfer Protocol) et fonctionne selon une architecture client/serveur. Les principes de l'hypertexte, qui permet à l'utilisateur de naviguer intuitivement à l'intérieur des documents par l'intermédiaire des liens, et de l'hypermedia, qui consiste à intégrer des documents de supports différents (texte, son, images fixes et animées), sont à l'origine de la popularité des serveurs Web.

Nous verrons par ailleurs que différents logiciels gravitent autour d'un serveur Web : logiciels serveurs, logiciels clients, convertisseurs ou utilitaires.

1.1. SCHÉMA DU SERVEUR DU CNS



LIAISON CNET GRENOBLE - INTERNET

* il permet de restreindre l'accès externe à un réseau privé spécifique

1.2...COMMENT FONCTIONNE UN SERVEUR WEB?

1.2.1 L'architecture client/serveur

Une application client/serveur est composée de deux programmes complémentaires tournant souvent sur des machines différentes. Le client exprime des requêtes et le serveur héberge la base de données, les deux communiquent selon un protocole. L'échange entre le serveur et le client est sans état c'est à dire qu'il n'y a pas de connexion permanente. Le client et le serveur n'occupent la ligne que durant le transfert des données.

Sur le serveur Web, un process détaché, le httpd (hypertext transfer protocol daemon) "écoute" le monde Internet sur un port TCP, le port 80 par défaut.

Les clients peuvent être sur n'importe quel ordinateur connecté sur Internet sous le protocole IP. Les serveurs peuvent être n'importe quel ordinateur d'un réseau sur lequel tourne un process détaché httpd. L'important est que ces réseaux puissent communiquer à l'aide du protocole TCP/IP.

Dans l'architecture client/serveur, le serveur Web distribue les documents répondant à une requête du client. Il est également chargé de la mise en forme des documents hypermedia.

Quand le client demande l'exécution d'un programme externe, c'est le serveur qui s'en charge. C'est en effet lui qui contient les programmes qui s'exécutent après le remplissage d'un formulaire ou après sélection d'un point dans une image réactive. Ces programmes se trouvent généralement dans un répertoire nommé cgi (common gateway interface).

Pour les documents écrits en langage HTML⁴, le client pourra exécuter des programmes de traitement des images réactives et de formulaires.

Quand le document contient une image ou du son, le client utilise un programme externe (xv sous Unix pour les images). Le client devient alors le serveur des ressources graphiques qui affichent le document. Toutefois, Netscape, logiciel client le plus utilisé actuellement, interprète sans programme extérieur les images au format GIF ou JPEG.

* il permet de restreindre l'accès externe à un réseau privé spécifique

⁴ HTML+ (ou HTML 3.0) est une version d'HTML en cours d'élaboration et déjà utilisable pour certaines fonctions, notamment la création de tableaux, l'intégration de formules mathématiques...

1.2.2 Le protocole HTTP

a. Fonctionnement d'HTTP

Il s'agit du protocole de transfert des documents hypertextes. Il fonctionne via le protocole TCP/IP et utilise une connexion par demande de document.

L'émission d'un document par un serveur HTTP vers un client Web se passe de la façon suivante:

- Le client effectue une connexion TCP sur le port HTTP du serveur ciblé, le serveur accepte la connexion;
- Le client émet sa requête, il s'agit d'une ligne de texte terminé par les caractères CR (Carriage Return) et LF (Line Feed). Le corps d'une requête est composé de la commande GET suivie d'un espace et de l'adresse du document souhaité;
- Le serveur expédie le document demandé, c'est un fichier HTML ;
- Le serveur coupe la connexion, elle matérialise aussi la fin du document.

b. L'adressage HTTP

La syntaxe est la suivante :

`http://nom_machine_serveur:port/nom_de_fichier`

http indique que le protocole HTTP est utilisé pour accéder à un serveur Web,

nom_machine_serveur est le nom du serveur Web au format Internet ou son adresse numérique (adresse IP),

Port est le numéro de port utilisé par le serveur pour recevoir la requête (80 par défaut).

L'adressage HTTP est un URL (Uniform Resource Locator).

1.2.3 Les URLs

World Wide Web utilise les Uniform Resource Locators (URL) pour définir l'adresse des documents sur d'autres serveurs. Les documents ainsi référencés peuvent être accessibles par différents protocoles (HTTP, FTP, GOPHER, WAIS...)

Ces documents peuvent être non seulement des fichiers mais aussi, par exemple, le résultat d'un programme ou de l'interrogation d'une base de données.

Le format d'un URL est :

`méthode://nom_de_machine:port/nom_de_fichier`

Le champ méthode indique le protocole à utiliser.

Le reste de l'URL dépend du champ méthode. En général, il consiste en un nom de machine, éventuellement suivi d'un double point et d'un numéro de port suivi d'un nom de fichier. Le numéro de port n'est à préciser que si le serveur n'utilise pas les valeurs standards.

Ainsi, l'accès au NCSA's Beginner's Guide to HTML se fera avec l'URL suivant :

`http://www.ncsa.uiuc.edu/General/Internet/WWW/HTML.Primer.html`

L' URL permet donc d'adresser sans ambiguïté toute ressource accessible sur Internet.

a.URL relatif

L'utilisation d'URLs relatifs permet de simplifier l'écriture des documents HTML.

Un URL relatif consiste soit en un nom complet d'un fichier, soit, en un nom relatif au document où l'on se trouve. Il faut entendre par le nom complet d'un fichier, son nom par rapport au répertoire racine du serveur Web.

On peut utiliser un URL relatif dans un document pour référencer un autre document localisé sur le même serveur et accessible par le même protocole. L'utilisation d'URLs relatifs permet de simplifier l'écriture des documents HTML. Il est conseillé d'utiliser le nom complet du document (relativement au répertoire racine du serveur), car celui-ci restera valide même si le document où il est utilisé change de place.

b.URL absolu

Un URL absolu est défini suivant le protocole d'accès aux fichiers existants. Ce protocole est indiqué par le champ méthode.

Par exemple dans le document

`http://user/limoux/Public/Mosaic/trillat/Formulaires/bibsimple.html`

nous avons un URL du type

`/trillat/Formulaires/infowais.html`

dont l'équivalent est en absolu:

`http:// user/limoux/Public/Mosaic/trillat/Formulaires/infowais.html`

1.3 LE LOGICIEL SERVEUR

Le logiciel serveur envoie les documents aux clients et exécute les programmes externes. Il existe différents logiciels serveurs et tous n'offrent pas les mêmes possibilités.

Dans notre cas, le serveur choisi est celui du CERN avec une plateforme Unix. Ses principales fonctionnalités sont les suivantes : autorisation d'accès, images cliquables, exécution de scripts par le serveur... Il supporte la version 1.3 de HTTP 1.0 et se compose d'un process détaché (le démon) et de quelques fichiers de configuration.

1.4 LE LOGICIEL CLIENT

Des logiciels clients Web (aussi appelés browser, navigateurs ou butineurs de réseau), existent pour pratiquement toutes les plateformes. Ils permettent de naviguer à travers Internet.

Grâce au client, l'utilisateur peut soumettre des requêtes au serveur Web de son choix et visualiser les réponses.

Le logiciel client peut également dialoguer avec d'autres types de serveurs (Wais, Archie, FTP, Gopher...).

Le logiciel client du CNS est Netscape version 1.1. Selon des statistiques élaborées par le serveur américain Web Random Yahoo Link⁵ en février 1995, Netscape est le client choisi dans plus de 77% des cas.

Pour afficher les documents HTML, Netscape utilise une mémoire cache. C'est dans cette mémoire qu'est chargé le document que l'on obtient en ouvrant un URL ou en cliquant sur un lien. Avant de charger un document en mémoire, Netscape vérifie si le document a été modifié. Mais, si au moment de tester les pages HTML d'un serveur, on utilise le bouton Back de la barre d'outils, Netscape présente la version stockée dans sa mémoire. Il faut donc enregistrer les modifications grâce au bouton Reload si l'on veut visualiser les derniers changements.

2. RÉALISATION DU SERVEUR

2.1... LES DOCUMENTS HTML

2.1.1 La norme SGML⁶

SGML (Standard Generalized Markup Language) est une norme d'échange de documents informatisés établie par l'ISO⁷ (norme ISO 8879).

La norme SGML a été adoptée par le projet CALS⁸. Elle intéresse toutes les entreprises, les administrations ou les maisons d'édition qui produisent de gros volumes d'information.

SGML permet de décrire un document générique mais cette norme s'intéresse uniquement à la structure logique de l'information contenue dans les documents.

Elle utilise un mécanisme de balisage compris à la fois par l'utilisateur et par l'ordinateur. La balise précédant une unité de texte donne toutes les indications nécessaires à la présentation de ce texte.

SGML définit ainsi des types de documents appelés DTD (Description Technique du Document).

La DTD sert de modèle à tous les documents d'un type de données. Le document qui respecte une DTD peut être retraité sur n'importe quelle plate-forme matérielle.

2.1.2 La DTD HTML

HTML (Hyper Text Markup Language) est une DTD de la norme ISO SGML. Elle a été conçue au CERN par Tim Berners Lee et Dan Connolly.

⁵ <http://www.cen.uiuc.edu/~ejk/bryl.html>

⁶ le serveur relatif à cette norme est à l'adresse suivante : <http://www.sil.org/sgml/sgml.html>

⁷ informations sur les normes et activités de l'ISO à <http://www.iso.ch/welcomef.html>

⁸ Computer-aided Acquisition and Logistic Support : programme établi par le Département de la défense américaine pour normaliser l'échange de documents à travers des systèmes hétérogènes.

Tout document balisé HTML peut être utilisé entre des plates-formes et des environnements différents.

HTML est composé d'un ensemble d'éléments appelés balises. Elles définissent la structure et la forme d'un document, la manière de l'afficher...

2.1.3 Les balises HTML

Les commandes HTML ont une balise de début et une balise de fin, toutefois, certaines marques de fin sont facultatives. Les caractères "<" et ">" permettent de délimiter les balises, celles-ci peuvent être écrites indifféremment en minuscules et en majuscules.

La liste qui suit n'est pas exhaustive, elle ne présente que les éléments essentiels formant le document HTML :

a. Les En-têtes

<HTML> </HTML> début et fin du document

<HEAD> </HEAD> informations non affichées concernant le document

<TITLE> </TITLE> titre du document affiché dans la barre de titre du client

<BODY> </BODY> Corps du document contenant toutes les informations qui seront affichées

<BASE HREF="URL"> </BASE> URL de base pour toutes les références contenues dans ce document.

Par exemple : <BASE HREF="user/limoux/Public/Mosaic/trillat">

Ce qui signifie que tous les objets appelés dans cette page seront dans des répertoires accessibles par ce chemin.

b. Les délimiteurs

Ils permettent d'aérer le texte car les clients Web n'interprètent pas les espaces et retours à la ligne.

<P> saut de paragraphe

 saut de ligne

<HR> insertion d'une ligne horizontale de séparation

Pour ces trois commandes, les marques de fin sont inutiles.

<DL> liste de description

<DT> 1er terme

<DD> description du 1er terme

</DD> fin de description du 1er terme

</DT> fin du 1er terme

</DL> fin de liste de description

c. Les styles

On peut faire varier la taille des caractères de H1 (très grande taille) à H6 (très petite taille).

<BLINK> </BLINK> permet de faire clignoter le texte compris entre les balises.

`<PRE>` `</PRE>` encadre un paragraphe formaté, celui-ci apparaît alors tel qu'il a été saisi.

Le texte peut également être mis en caractères gras `<BOLD>` `</BOLD>` ou en italique `<I>` `</I>`.

d. Les Caractères accentués et les symboles

Un document HTML est écrit en code ASCII 7 bits, ce qui pose problème pour les caractères accentués.

La représentation des caractères nationaux est donc assurée par des séquences d'échappement qui s'utilisent comme suit:

`î` pour î

`ç` pour ç

`é` pour é

`à` pour à ...

Les caractères `<` `>` et `&` étant interprétés par HTML, ils doivent être représentés ainsi :

`<` pour `<`

`>` pour `>`

`&` pour `&`

e. Les listes

Il en existe deux sortes :

numérotée :

``

`` 1er élément (`` est inutile)

`` 2ème élément

``

ou non numérotée :

``

`` 1er élément

``

f. Les images dans le texte

HTML permet l'insertion d'images dans un texte. Celles-ci peuvent apparaître dans le corps du texte mais également servir d'ancres. (la notion d'ancre est développée plus loin).

La commande `` permet d'insérer des images dans le texte, sa syntaxe est la suivante:

`` pour insérer une image distante ou

`` pour insérer une image locale.

Nom_fichier_local étant un chemin relatif ou absolu dans l'espace accessible au serveur Web correspondant.

On peut ensuite positionner l'image au centre grâce à l'attribut `<CENTER>` ou l'aligner à droite ou à gauche :

``

``

g. les images réactives

Une image réactive permet de faire correspondre des documents à des zones déterminées de cette image. Suivant l'endroit de l'image où l'on clique, tel ou tel document sera chargé et interprété par le client. Ainsi, la carte de France réactive⁹ du CNRS permet d'afficher la carte de la région sur laquelle on clique. Les images réactives sont entourées d'un liseré bleu afin qu'on les distingue des images non réactives.

La syntaxe pour décrire une image est la suivante :

```
<METHODE><URL>coord1 coord2 coord3...coordn
```

Coord sont les coordonnées au format x,y. La valeur de ces coordonnées est le nombre de pixels en x et y, l'origine étant le point en haut à gauche de l'image réactive.

Le nombre de coordonnées dépend de la méthode qui peut être circle, poly ou rectangle.

Un cercle est défini par deux points : le premier est son centre, le second un point de son périmètre.

un polygone peut avoir jusqu'à cent sommets, ils sont définis par leurs coordonnées en x et en y.

Un rectangle est défini par les coordonnées de ses sommets en haut à gauche et en bas à droite.

URL est l'adresse du fichier qui va être envoyé.

Pour calculer les coordonnées des images, le plus simple est d'utiliser le logiciel xv qui permet de visualiser les images tout en donnant les coordonnées du point sur lequel se trouve le curseur de la souris.

Ultérieurement, car ce n'est pas une priorité, nous créerons grâce à cette commande, un plan de la bibliothèque cliquable en divers endroits.

Imagemap est un exécutable qui met en oeuvre les images réactives. On le place généralement sous le répertoire cgi-bin dans lequel se trouvent tous les programmes binaires utilisés par le serveur. Quant au fichier de configuration des images réactives (fichier imagemap.conf) il est rangé dans le répertoire conf, avec toutes les images concernées (au format gif).

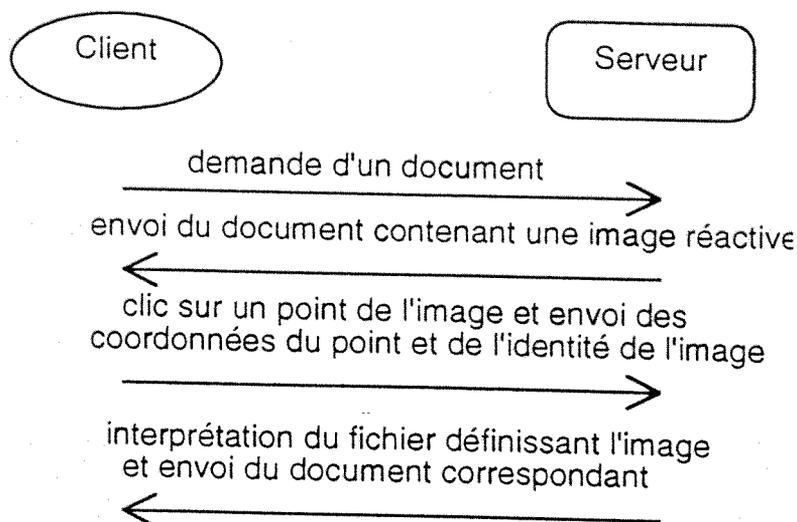
Côté client, le document HTML est interprété et affiche l'image réactive définie par le fichier gif.

L'utilisateur clique sur l'image, la directive ISMAP envoie au serveur les coordonnées du point de la carte où le curseur se trouvait au moment du clic, ainsi que le paramètre (carte dans notre exemple) associé à l'image réactive :

```
<A HREF="/cgi-bin/imagemap/carte"> <IMG SRC="images/exemple-carte.gif" ISMAP> </A>
```

Le déroulement des échanges, lorsqu'il y a une image réactive peut être résumé ainsi :

⁹ <http://web.urec.fr/france/france.html>



h. l'attribut background

Cet attribut permet d'incorporer une image en transparence sur le fonds d'une page HTML.

Il s'utilise comme suit :

```
<BODY BACKGROUND= "URL de l'image">
</BODY>
```

Il convient toutefois d'en faire un usage modéré. En effet, la présence d'un arrière-plan augmente le temps d'affichage de la page et certains fonds rendent le texte difficile à lire. Nous avons également constaté que sur les écrans des terminaux X, le fonds n'apparaît que partiellement, le résultat esthétique est alors à l'opposé de l'effet recherché.

Pour toutes ces raisons, nous avons préféré ne pas utiliser cet attribut, chaque page contenant suffisamment de texte pour ne pas être alourdie par l'ajout de couleurs ou de dessins.

i. Les liens hypertextes

La puissance d'HTML vient de la possibilité d'établir un lien entre des mots d'un document (ou d'une image) vers un autre document.

Le client met en évidence ces mots (généralement par une couleur et/ou un soulignement) pour indiquer qu'il s'agit d'un lien hypertexte.

La balise de l'hypertexte: <A> est aussi appelée ancre.

Pour insérer une ancre dans un document il faut :

- * Signaler le début de l'ancre par <A>
- * Indiquer le *document qui sera lié* avec le paramètre
- * <HREF="document">
- * Saisir le texte qui servira de **lien hypertexte**
- * Terminer la balise de l'ancre par

ce qui donne :

- lien entre des documents situés dans un même répertoire

 Retour à la première page

La phase soulignée est le lien hypertexte vers le document presentation.html, lequel est situé dans le même répertoire que le document initial.

- lien entre des documents situés dans des répertoires différents

Pour cela, il faut préciser le chemin relatif vers le document d'arrivée. Les liens relatifs sont préférables car ils sont plus courts à écrire (ce qui réduit le risque d'erreur) et si l'ensemble des documents est déplacé, les liens hypertexte restent valides.

Les chemins utilisent la syntaxe d'UNIX.

 Retour à la première page

- lien vers un paragraphe d'un autre document

Une ancre peut faire référence à un paragraphe d'un autre document HTML. Pour cela, il faut en premier, marquer le paragraphe du document qui sera lié (mot-clé NAME) :

<NAME = "explication"> explications sur le mode d'interrogation ...

Il faut ensuite créer dans le document d'origine le lien vers le paragraphe, en y incluant le nom du document et la marque du paragraphe séparé par le caractère "#".

 lien vers le paragraphe d'explications sur le mode d'interrogation.

- lien à l'intérieur d'un document

La méthode est identique à celle décrite précédemment mais sans indication du nom du document.

L'utilisation du caractère "#" suivi d'une chaîne de caractères à la fin du nom d'un document permet de référencer un endroit (une ancre) dans celui-ci.

< A HREF = " #explication" > lien vers le paragraphe d'explications sur le mode d'interrogation.

j. Appel d'un programme

Toutes les informations accessibles sur le réseau Web n'existent pas uniquement sous forme de fichiers.

Il est également possible d'exécuter des programmes (aussi appelés scripts), pour cela :

- le client envoie sa requête au serveur
- le programme est exécuté

- la réponse est envoyée au serveur
- celle-ci est ensuite redirigée vers le client qui l'affiche.

Si le programme contient la variable Content-Type=HTML, la réponse peut être visualisée directement sur le serveur car il s'agit d'une page HTML.

Un programme exécutable sur le Web peut être écrit dans n'importe quel langage ou système d'exploitation.

Le programme doit comprendre au moins ces trois étapes :

1. Récupération des paramètres passés au programme et leur interprétation.
2. Réalisation de l'objet même du programme à partir des paramètres.
3. Affichage du résultat sur la sortie standard.

Cette sortie est automatiquement redirigée vers le client qui a effectué la requête.

Pour exécuter un programme, il faut des commandes dans un document HTML. Cela peut se faire par un lien direct ou bien par un formulaire.

* exécution du programme par un lien

Pour créer un lien vers un programme il suffit d'inclure une ligne de commande de ce type dans un document HTML:

```
<A HREF=adresse_du_programme/nom_du_programme? paramètre1
=valeur1 & paramètre2=valeur2> exécute le programme</A>
```

Cette commande crée un lien classique : le fait de cliquer sur les mots "exécute le programme" fait sauter le client vers un autre document. La seule différence est que le document affiché sera le résultat d'exécution du programme.

L'adresse_du_programme est son adresse au sens URL, à savoir, soit le chemin d'accès dans le serveur courant, soit une adresse URL complète.

Exemple d'appel qui permet de faire afficher le résultat de la requête Wais sur notre serveur:

```
<A HREF=/Public/Mosaic/htbin/bib3.sh?pid=1229&numero=1> Un
serveur W3 au CNET de Grenoble</A>
```

Si l'utilisateur clique sur la réponse à sa requête, ici le titre de la notice, il exécutera le programme appelé bib3.sh. Ce programme se trouve sur le serveur Web. Il est exécutable et placé dans le répertoire "ht/bin" sous la racine du serveur. Les paramètres sont : PID¹⁰ qui se

¹⁰ PID (Process Identifier) terme Unix attribuant un numéro séquentiel à un processus

voit affecté la valeur "1229" et numero qui prend la valeur "1". L'utilisateur recevra une page contenant le résultat en sortie du programme, ici la notice bibliographique complète.

Pour passer des arguments, il est également possible d'utiliser un formulaire. Le formulaire est une fiche que remplit l'utilisateur.

* exécution du programme par un formulaire

La création de formulaires dans un document HTML se fait par l'utilisation de la commande suivante :

```
<FORM>  
<FORM ACTION="nom_du_programme"[METHODE=mot_clef]>  
</FORM>
```

Action est un programme exécutable ou une procédure (shell, perl, ...).

Mot_clef est la méthode : soit POST, soit GET.

Avec la méthode POST les paramètres sont envoyés sur l'entrée standard du programme. Cette méthode est recommandée quand le nombre de paramètres est important pour éviter que la ligne de commande ne soit tronquée (Unix accepte 255 caractères au maximum dans une chaîne).

Dans le cas de la méthode GET, le nom du programme est suivi d'un point d'interrogation qui indique qu'il est suivi d'un ou plusieurs paramètres. Les paramètres sont séparés par le signe &. Un paramètre est constitué de son nom, suivi du signe = et de la valeur qui lui est affectée. Si cette valeur comporte plusieurs mots les espaces entre ces mots sont remplacés par des +.

Les variables sont transmises derrière l'URL sous la forme d'une phrase :

```
nom_de_programme?paramètre1=valeur1&paramètre2=valeur2&param  
ètre3=valeur3...
```

Les données sont facilement extractibles car regroupées dans un fichier dont l'adresse est indiquée par la variable FILE en en-tête du programme script.

Les informations extraites de la base de données étant affichées à l'écran sous forme de pages HTML, il faut utiliser des commandes HTML par l'intermédiaire de la commande Unix echo.

Ensuite, la variable QUERY-STRING est utilisée pour rechercher des informations dans les bases de données.

Enfin, dans tout formulaire, il faut obligatoirement placer un bouton de type SUBMIT. C'est par lui que l'utilisateur soumet sa requête. On y adjoint généralement un bouton de type RESET qui permet de remettre les champs à vide.

Comme pour l'exécution par un lien, le programme doit être exécutable et placé dans un répertoire (user/limoux/Public/Mosaic/scripts) autorisé en exécution par le serveur sous lequel il est placé.

Le résultat affiché peut être du texte simple, il est cependant intéressant de formater le texte suivant la syntaxe HTML. Il sera interprété par le client comme un document HTML normal.

La page suivante présente le formulaire abonnements.html permettant de retrouver le titre d'un périodique. L'utilisateur peut proposer jusqu'à trois termes.

ABONNEMENTS

● Les périodiques sont à consulter sur place.

Ceux de l'année en cours sont exposés au fond de la bibliothèque et classés par ordre alphabétique de titre.

Les années antérieures se trouvent en archives. L'état des collections ainsi que leurs localisations sont affichées à l'entrée des archives.

Pour savoir si la bibliothèque est (ou a été) abonnée à celui qui vous intéresse, saisissez un ou plusieurs mots *significatifs* de son titre.

▲ Entrez un terme ou le début d'un terme seulement par case. (3 lettres minimum, pas d'article, ni accent)

Si le nom du périodique comprend une abréviation, entrez la sans espaces. (Ex : CAE pour Computer-Aided Engineering.)

IL n'est pas possible de différencier IEE de IEEE, donc si vous n'interrogez que sur ce terme, vous obtiendrez la liste confondant les deux séries.

Vous pouvez ne remplir que la première ou que les deux premières cases mais pas la 2ème ou la 3ème seule ni les deux dernières seules.

(Ex : *journal* dans la première case *telephone* dans la seconde pour *Le journal du téléphone.*)

ancer la recherche	facier les champs

Liste des abonnements.



Retour au menu principal

```

<TITLE>liste des abonnements</TITLE>
<CENTER>
<IMG SRC="/trillat/Icons/nouveau.gif">
<H3> ABONNEMENTS </H3>
</CENTER><IMG SRC="/trillat/Icons/blueline.gif">
<BR><BR>
<FORM ACTION="http://161.106.20.51/htbin/abonnements.sh"
METHOD="GET">
<H3> <IMG SRC="/trillat/Icons/ballerouge.gif"> Les périodiques
sont &agrave; consulter sur place.
<BR><BR>
Ceux de l'année en cours sont exposés au fond de la bibliothèque et
classés par ordre alphabétique de titre.
<BR><BR>
Les années antérieures se trouvent en archives. L'état des
collections ainsi que leurs localisations sont affichés à l'entrée des
archives.
<BR><BR>
Pour savoir si la bibliothèque est (ou a été) abonnée à celui qui vous
intéresse, saisissez un ou plusieurs mots <I>significatifs</I> de son titre.
<BR><BR>
<IMG SRC="/trillat/Icons/warning.gif"> Entrez un terme ou le début d'un
terme seulement par case.
<BR><BR>
Vous pouvez ne remplir que la première ou que les deux premières
cases.
<BR><BR>
(Ex : <I> journal</I> dans la première case <I> telephone </I> dans la
seconde pour <I> Le journal du téléphone. </I>)
<BR><BR>
<CENTER>
<INPUT NAME="mot1" SIZE=35>
<BR>
<INPUT NAME="mot2" SIZE=35>
<BR>
<INPUT NAME="mot3" SIZE=35>
<BR>
<INPUT TYPE="submit" VALUE="Lancer la recherche ">
<INPUT TYPE="reset" VALUE="Effacer les champs" >
<BR><BR>
</CENTER>

```

Ce formulaire est ensuite traité par le programme script abonnements.sh

```
#!/bin/sh
FILE="/user/limoux/Public/Mosaic/trillat/Pages/alphaabonment.html"
export QUERY_STRING
echo "Content-Type: text/html"
eval `/user/limoux/Public/Mosaic/scripts/cgiparse -form`
echo
echo '<PRE><H3>'

FORM_mot1=${FORM_mot1:=1}
FORM_mot2=${FORM_mot2:=1}
FORM_mot3=${FORM_mot3:=1}

#aucun critere
[ $FORM_mot1 = "1" ] [ $FORM_mot2 = "1" ] [ $FORM_mot3 = "1" ] &&
echo "Aucun critere" && exit

#tous les criteres
if [ $FORM_mot1 != "1" -a $FORM_mot2 != "1" -a $FORM_mot3 != "1" ];
then
echo "voici le résultat de votre interrogation : "
grep -i "$FORM_mot1" $FILE | grep -i "$FORM_mot2" | grep -i
"$FORM_mot3"
[ $? -eq 1 ] && echo "aucune reponse avec ces criteres "\
&& exit
fi

#critere1 uniquement
if [ $FORM_mot1 != "1" -a $FORM_mot2 = "1" -a $FORM_mot3 = "1" ]; then
echo "voici le résultat de votre interrogation (mot 1) : "
grep -i "$FORM_mot1" $FILE
[ $? -eq 1 ] && echo "aucune reponse avec le 1er critere de recherche" \
&& exit
fi

#criteres 1 et 2 uniquement
if [ $FORM_mot1 != "1" -a $FORM_mot2 != "1" -a $FORM_mot3 = "1" ];
then
echo "voici le résultat de votre interrogation (mots 1 et 2) : "
grep -i "$FORM_mot1" $FILE | grep -i "$FORM_mot2"
[ $? -eq 1 ] && echo "aucune reponse avec les 1er et 2eme criteres de
recherche" \
&& exit
fi
```

Chaque champ est défini par la commande "INPUT" qui permet de préciser les attributs de celui-ci. Le formulaire offre à l'utilisateur quatre types de champs possibles : des boutons, des lignes de texte, des zones de texte et des menus, identiques à ceux des interfaces de types X- Windows ou Windows :

• **Bouton :**

<INPUT TYPE NAME VALUE CHECKED>

avec

-TYPE qui peut prendre différentes valeurs : radio (boutons multiples) ou checkbox (cases à cocher) ou submit (pour envoyer le formulaire une fois rempli) ou reset (pour remettre à vide tous les champs du formulaire).

-NAME est le nom qui permet d'identifier le bouton. Il sera transmis comme argument à une procédure chargée de traiter le contenu du formulaire.

-VALUE est la valeur donnée au bouton lorsqu'il est sélectionné

-CHECKED signale un bouton par défaut ou une case cochée avant saisie

• **Texte libre sur une ligne :**

<INPUT TYPE NAME SIZE>

avec

-TYPE peut prendre la valeur text (pour permettre la saisie d'un ou plusieurs mots) ou password (les caractères seront alors représentés par des *) ou hidden (le champ est caché)

- NAME=nom qui permet d'identifier le champ texte

• **Texte libre sur plusieurs lignes :**

<TEXTAREA>NAME ROWS COLS</TEXTAREA>

avec

- NAME = nom qui permet d'identifier le champ texte

- ROWS = nombre de lignes

- COLS = nombre de colonnes

• **Menu :**

- <SELECT NAME> nom qui permet d'identifier le menu

<OPTION SELECTED> représente le choix par défaut

<OPTION> représente chaque choix du menu

</SELECT>

Le formulaire de suggestion d'achat ainsi que son script de traitement sont présentés en annexe III, page 56.

2.1.4 Quelques règles d'ergonomie¹¹

La présentation de l'information est primordiale car elle permet au lecteur d'évaluer la qualité du serveur et sa fiabilité.

Le but de Web est de fournir une information utile à l'utilisateur qui doit accéder facilement à l'information qu'il recherche. Il faut donc structurer l'information (pages d'accueil et liens entre les documents) en essayant de se mettre à la place du lecteur et en tenant compte du type de public auquel on s'adresse.

Dans notre cas, les pages réalisées devaient s'intégrer harmonieusement à celles du serveur du CNS.

Les règles à appliquer sont donc les suivantes : simplicité, clarté, concision, textes aérés avec des liens ni trop fréquents, ni trop rares.

En effet, si les pages sont trop longues, le temps de réponse d'un document à l'autre sera important et l'utilisateur devra sans cesse faire fonctionner la barre de défilement.

A l'inverse, si les pages sont trop courtes, l'utilisateur sera sans cesse amené à cliquer pour passer d'une fenêtre à l'autre.

Il faut également éviter l'emploi excessif d'images qui peut allonger considérablement les temps de réponses.

Le titre donné comme ancre sur la page de présentation est systématiquement repris en en-tête de la page concernée ce qui permet de se repérer plus facilement.

Au bas de chaque page, une flèche ancrée permet de revenir à la page de présentation du serveur pour éviter à l'utilisateur de repasser par toutes les pages qu'il a déjà visionnées.

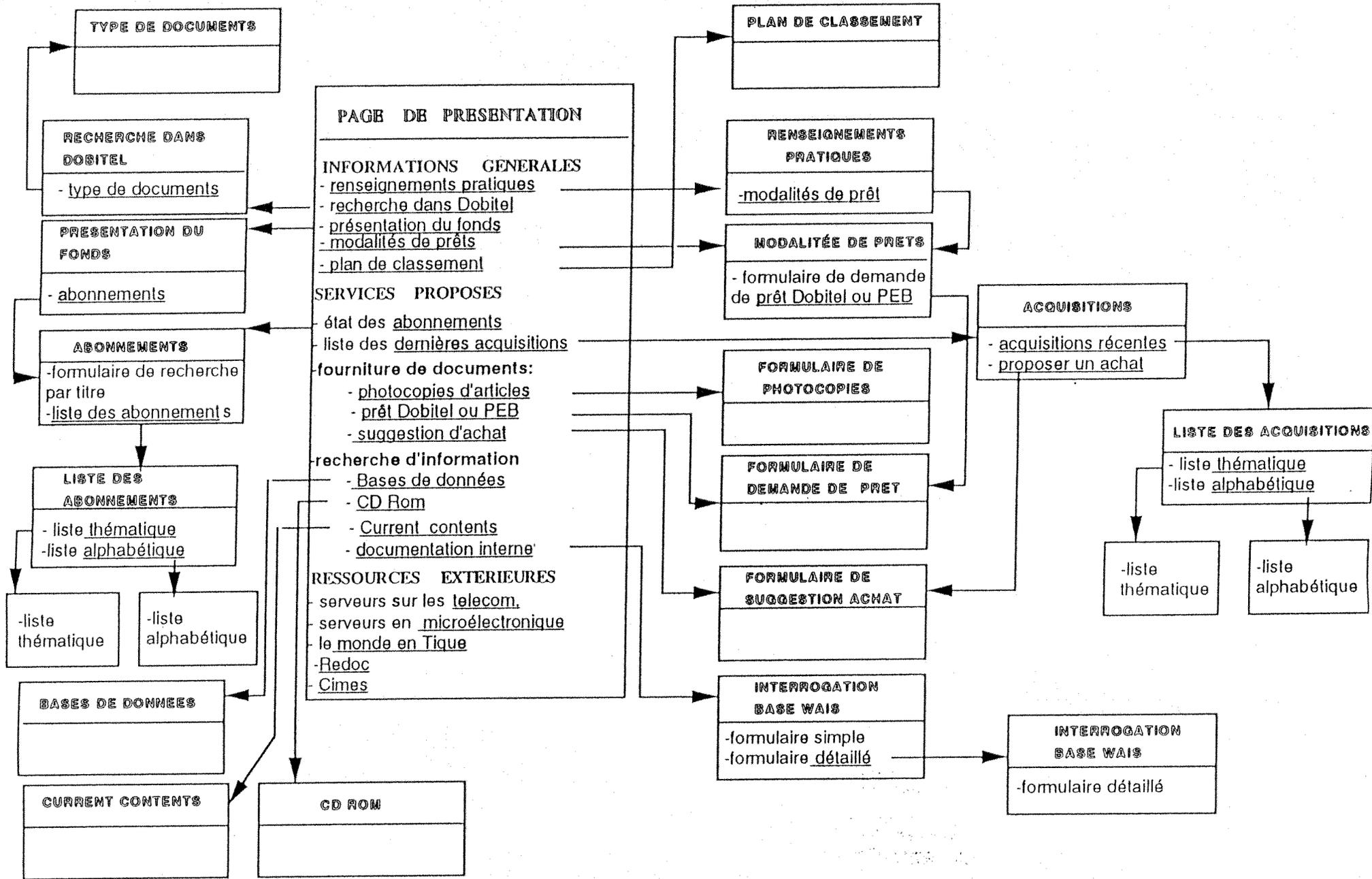
De même, sur cette page de présentation de la bibliothèque, une ancre permet de revenir au menu principal du serveur du CNS.

Enfin, il est important de tester le serveur sous différents clients et sur différents postes car des disparités de présentation apparaissent parfois.

Le schéma qui suit présente l'ensemble des pages existantes et les liens créés entre celles-ci.

¹¹ <http://www.w3.org/hypertext/WWW/Provider/Style/Overview.html>

La diffusion de ce document est soumise à autorisation du CNET



ORGANIGRAMME DES PAGES DU SERVEUR WEB

2.2...LES Outils DE RÉDACTION

Un document HTML est un document au format texte (ASCII). Il peut être écrit avec n'importe quel éditeur de texte mais il est préférable de recourir à l'un des logiciels dédiés à la rédaction de documents HTML. On distingue trois catégories d'outils de rédaction exploitables sur différentes plates-formes :

2.2.1 Les éditeurs HTML¹²

Ce sont des logiciels fournissant des balises HTML prédéfinies, accessibles à partir de barres d'outils. L'inconvénient est que très peu d'entre eux fonctionnent en environnement WYSIWIG¹³. On ne peut donc pas visualiser directement le résultat, il faut préalablement charger le document sous un logiciel client.

Après toute modification sous éditeur (et sauvegarde) l'option Reload permet de visualiser le changement.

Emacs est l'éditeur que nous avons utilisé. Il possède une variété assez importante de fonctions telles que l'édition simultanée de différents fichiers dans plusieurs tampons ou la pré-génération de lignes de commandes.

2.2.2 Les gabarits HTML pour Word

Ils permettent d'exploiter un document HTML sous le traitement de texte Winword 2.0 ou 6.0 dans un environnement WYSIWYG. Ils proposent généralement des commandes HTML prédéfinies sur une barre d'outils.

2.2.3 Les convertisseurs

Egalement appelés filtres¹⁴, ils permettent d'une part la récupération de documents existants dans un format différent, d'autre part l'utilisation de logiciels de traitement de textes évolués pour écrire un document destiné à être mis en ligne sur un serveur Web.

Nous avons ainsi utilisé RTFtoHTML¹⁵ pour intégrer la liste des périodiques auxquels la bibliothèque est abonnée.

Ce logiciel peut s'installer sur Unix ou sur Mac.

Une fois installé, on peut personnaliser la conversion grâce au fichier HTML-trans qui contient un ensemble de règles de conversion entre les styles de paragraphes et de caractères RTF vers ceux de HTML.

¹² <http://www.w3.org/hypertext/WWW/Tools/Overview.html>

¹³ WYSIWYG : What You See Is What You Get

¹⁴ le serveur du Cern propose, à l'adresse suivante, une liste des filtres existants:
http://www.w3.org/hypertext/WWW/Tools/Word_proc_filters.html

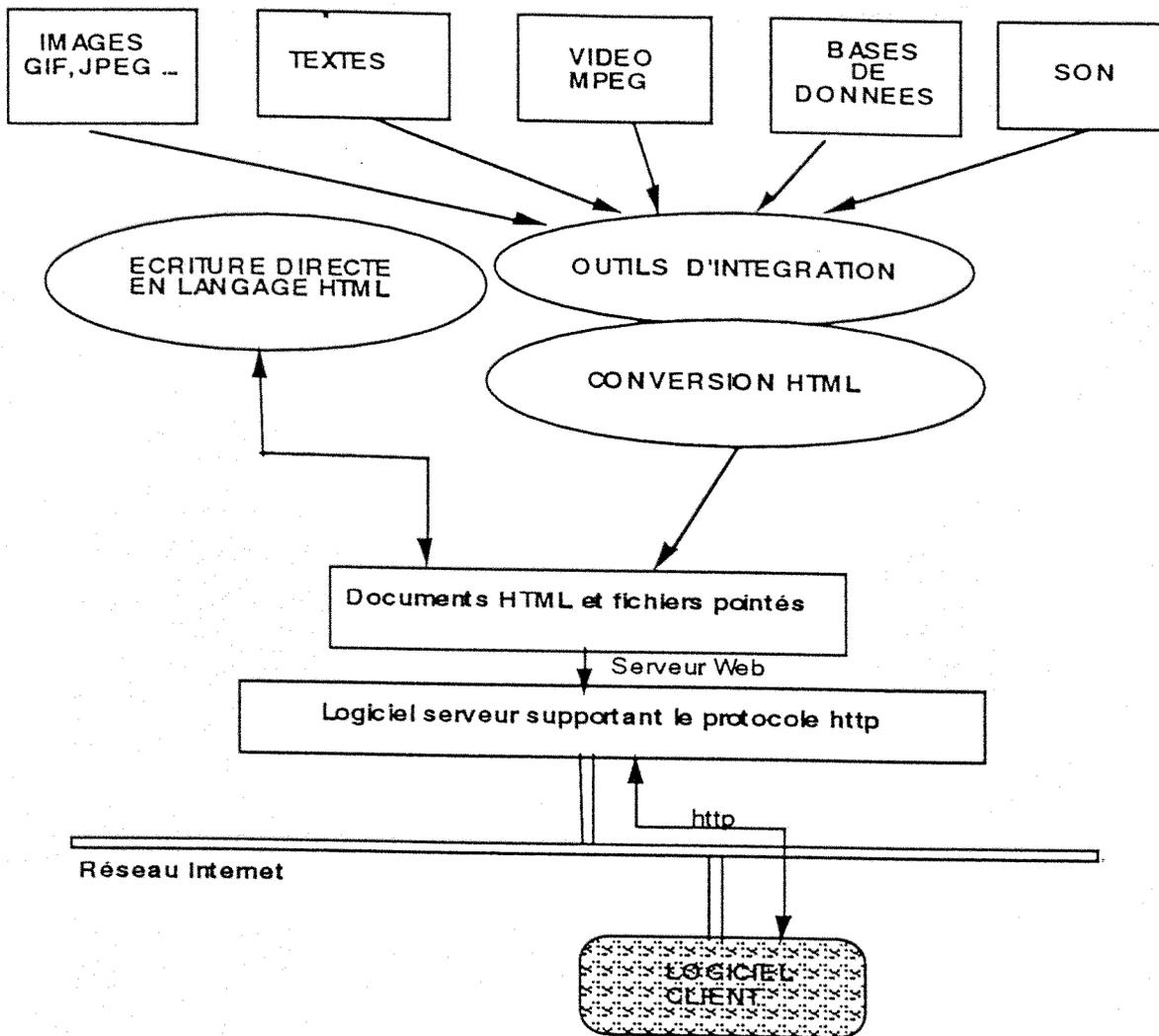
¹⁵ <http://www.w3.org/hypertext/WWW/Tools/rtftohtml-2.6.html>

2.3 L'INSERTION DES IMAGES FIXES

Que ce soit pour créer des images pour un serveur ou les faire visualiser par un client Web, il est nécessaire d'avoir un certain nombre de logiciels. Dans notre cas, il s'agissait uniquement d'insérer des icônes. Nous avons utilisé le logiciel xv qui fonctionne sous Unix, permet de visualiser des fichiers gif, jpeg, tiff, pbm, pgm, ppm, X11 bitmap, bmp, pcx. Il permet la conversion d'un format à l'autre et le changement de taille des images.

Le schéma de la page suivante décrit l'organisation des différents éléments composant un serveur Web.

ELEMENTS DE DEVELOPPEMENT
D'UN SERVEUR WEB



IV . LA BASE WAIS

1. PRÉSENTATION GÉNÉRALE

WAIS (Wide Area Information Server) est un ensemble de logiciels développés par Thinking Machines Corporation (TMC), Dow Jones et Apple Computer.

Ces logiciels permettent d'indexer des documents de divers types localisés sur des serveurs distants et d'interroger les bases existantes.

A l'origine, le projet Wais est une expérience de recherche et de collecte informatisée de documents hétérogènes. Le succès de ce projet repose sur la mise en place d'un grand nombre de serveurs d'information.

Afin d'aider au développement de ce marché, les spécifications et le code source sont diffusés gratuitement, mais les dernières versions ne sont plus développées par TMC.

La société WAIS Inc. propose des versions commerciales tandis que le CNIDR (Clearinghouse for Networked Information Discovery and Retrieval) poursuit le développement des versions du domaine public, les versions freeWAIS.

L'université de Dortmund a développé une version dite FreeWAIS-sf (search fields) qui ajoute aux fonctions antérieures (opérateurs booléens, troncature), la possibilité d'indexer des champs définis et non plus la totalité du document. C'est donc la version freeWAIS-0.3-sf1.2¹⁶ que nous avons choisie d'installer.

1.1...COMMENT FONCTIONNE UNE BASE WAIS?

Après installation, le serveur Wais se compose de deux parties : waisserver (le "démon") et waisindex (l'indexeur). Il faut également choisir un logiciel client parmi ceux disponibles : xwais, swais, waissearch, waisq (qui est la version que nous avons installée).

L'installation du serveur consiste à indexer les fichiers puis à lancer le serveur.

1.1.1 L'architecture client/serveur

Wais fonctionne sur le mode client/serveur et s'appuie sur le protocole Z39.50.

Grâce au modèle client/serveur, les problèmes du nombre d'accès simultanés à la machine ne se posent pas. Un serveur accepte les requêtes qui arrivent par le réseau, les traite et renvoie les réponses correspondantes.

1.1.2 Le protocole Z39.50

Il sert à formaliser les échanges entre une machine cliente et une machine serveur.

¹⁶ disponible en FTP anonyme à l'adresse suivante:
ls6-www.informatik.uni-dortmund.de

Z39.50 est un standard national aux Etats-Unis (ANSI/NISO Z39.50) destiné à l'échange d'information documentaire entre ordinateurs.

Il permet de rendre la recherche transparente à l'utilisateur dans des catalogues de bibliothèques en ligne (OPAC) et des sources d'information distantes. En effet, les systèmes de catalogue en lignes sont rarement conviviaux, à la différence de Wais, plus facile à utiliser.

1.2 LE LOGICIEL D'INDEXATION : WAISINDEX

Il est possible d'indexer des types très variés de documents : notices bibliographiques, archives de courriers électroniques, archives de news, textes intégraux aux formats ASCII, PostScript, MIME, SGML, ODA...

Les index de Wais contiennent toujours du texte, cependant on peut créer des bases contenant d'autres documents (images, sons, documents électroniques) en indexant les noms des fichiers ou de petits descriptifs.

A l'origine, Wais indexait le texte intégral des documents mais la version SF de l'université de Dortmund permet de distinguer les champs.

Pour indexer une base sur ses champs, il faut la structurer afin que ceux-ci soient facilement discernables. On décrit ensuite cette structure dans un fichier appelé <nom_de_base>.fmt.

(voir annexe IV, p.58)

Lors de cette description, on précise le séparateur de notices (le saut de ligne dans notre cas) la ligne d'en-tête¹⁷, l'index global (il sera inutile de le nommer pour l'interroger) et les index locaux (il faudra préciser le champ pour l'interroger). Ce fichier fmt est donc essentiel pour l'interrogation.

Une fois ce fichier créé, on exécute la commande waisindex avec les arguments -t fields (type du fichier) et -export si on souhaite rendre la base accessible de l'extérieur, par défaut elle est locale à la machine hôte.

Le logiciel d'indexation balaye le fichier contenant les documents à indexer, construit les index, des fichiers inverses pour chacun des champs définis et d'autres fichiers (.cat, .src, .dlm...).

1.3 LE LOGICIEL SERVEUR : WAISSERVER

Waisserver reçoit les requêtes des clients, interroge les bases choisies par l'utilisateur et renvoie les réponses classées selon leur pertinence.

Après copie de l'exécutable waisserver dans le répertoire où résident les bases, on choisit le port sur lequel on veut le faire tourner (210 par défaut).

On peut ensuite le lancer en "standalone" ou par "inted".

Dans le premier cas, il tournera tant qu'on ne l'arrête pas et il faut inclure une commande pour le relancer chaque fois que l'on redémarre la machine, c'est l'option que nous avons choisie.

¹⁷ la ligne d'en tête est appelée headerline, c'est ce que voit l'utilisateur en réponse à une question

Dans le second, le serveur ne tourne que s'il est sollicité par la requête d'un client.

1.4 LE LOGICIEL CLIENT : WAISQ

La partie client du logiciel permet à l'utilisateur d'exprimer des requêtes et affiche les réponses du serveur. Des clients existent pour un grand nombre de machines (Unix, Macintosh, PC).

Les fonctionnalités sont les suivantes :

Nous avons vu que l'utilisation du caractère "?" suivi d'une chaîne de caractères à la fin du nom d'un document correspond à une liste de paramètres pour l'exécution d'un programme. Mais il peut aussi s'agir de l'interrogation d'un document indexé dans une base Wais. Dans ce cas, les caractères spéciaux (blancs, caractères accentués....) sont codés. Par exemple:

`wais://quake.think.com:210/directory-of-servers?inria`

correspond à l'interrogation de la base WAIS directory-of-servers sur le serveur *quake.think.com*. *inria* étant le paramètre de l'interrogation.

1.5 LE NIVEAU DE DIFFUSION DE LA BASE

Wais ne prévoit pas d'accès réservé à la connexion mais il est possible de lui attribuer différents niveaux de diffusion : base enregistrée dans l'annuaire mondial; dans l'annuaire national (à l'étranger, seuls les utilisateurs sachant qu'elle existe et connaissant ses caractéristiques peuvent alors l'interroger); base exportée (non enregistrée dans les annuaires publics) ou base locale à la machine hôte.

Si la base n'est déclarée nulle part, elle garde un caractère privé. Toutefois, cela n'exclue pas l'accès depuis une machine extérieure. Pour plus de sécurité, L'INRIA propose de faire tourner le serveur sur un port autre que 210 (habituellement dévolu à Wais), c'est la solution que nous avons choisie, de plus, l'accès n'est autorisé qu'aux adresses IP des autres Cnet et Ecoles des télécommunications grâce à la présence de firewalls. (voir schéma p.21)

Car s'il est envisageable que le serveur Web soit ouvert un jour au public extérieur, les notes techniques ne seront en revanche jamais rendues publiques en raison de leur degré de confidentialité.

De même, les articles en voie de publication (pré-publications), ne pourront être disponibles car leur diffusion appartient à ce moment là à la maison d'édition ou au journal concerné.

Seuls les articles déjà parus seront éventuellement accessibles de l'extérieur.

2. REALISATION DE LA BASE

2.1 OBJECTIFS DE LA BASE

Depuis quelques mois, la direction a décidé d'inclure les rapports de stage dans la catégorie des notes techniques, ceci afin de mieux gérer leur conservation et de les mettre plus facilement à la disposition des lecteurs potentiels.

2.2... STOCKAGE DE LA BASE

Par ailleurs, aucun document antérieur à 1995 ne sera archivé dans cette base et l'accroissement prévu est d'environ deux documents par semaine, soit une centaine de notes techniques par an. Mon travail consiste donc à intégrer le maximum de notes de l'année en cours.

A terme, l'objectif est de créer deux autres bases Wais dans lesquelles seront stockées les thèses et les publications de chercheurs. Cette dernière base devant être présentée comme la vitrine du CNS le jour où le serveur deviendra accessible de l'extérieur.

2.3... MODE D'INTERROGATION

L'interrogation renvoie les titres des documents pertinents qui sont sous forme d'ancres cliquables. Ces ancres renvoient à la notice complète du document choisi. Dans cette notice, une autre ancre propose à l'utilisateur de consulter le document dans son intégralité. Pour cela, il faut donc récupérer dans chaque secrétariat correspondant, le document sur support électronique, si toutefois il a été conservé. Il faut ensuite traduire ce document Word, comportant des graphiques et schémas le plus souvent réalisés sous MacDraw, en un document HTML, lisible depuis le serveur.

2.4... PROBLEMES D'INTÉGRATION DE DOCUMENTS COMPLETS

Un problème est apparu au moment de cette traduction. En effet, lorsqu'on enregistre un fichier Word au format RTF, les schémas sont enregistrés au format pict. Ils ne peuvent donc pas être traduits ensuite par le convertisseur RTFtoHTML.

La solution aurait été d'aller chercher chaque graphique dans le document Word, de l'enregistrer au format gif puis de le transférer dans document HTML en ajoutant une ancre. Ce travail, assez fastidieux, mais envisageable si le nombre de dessins est réduit, s'avère difficilement réalisable dans notre cas.

En effet, s'il est sans doute possible d'automatiser au moyen d'un script le passage d'un format à l'autre, resterait ensuite à prospecter manuellement dans le document afin de récupérer chaque graphique.

Il existe pour Mac et PC divers produits permettant vraisemblablement de régler ce problème mais notre serveur tournant sous Unix, il serait préférable de trouver un outil correspondant. Une tentative avec le logiciel Tgif n'a pas donné les résultats escomptés, nous continuons donc de prospecter sur Internet.

Nous nous orientons vraisemblablement vers la solution suivante : le document complet sera stocké au format Postscript. Quand

l'utilisateur cliquera sur l'ancre pour récupérer le texte intégral, le fichier sera enregistré dans son répertoire de travail. Cela suppose que les utilitaires correspondants, à savoir Ghostscript et Ghostview, soient mis à disposition sur le réseau.

2.5...PASSERELLE WAIS/ WEB¹⁸

Comme nous l'avons décrit plus haut, l'interrogation de la base Wais se fait par l'intermédiaire d'un formulaire depuis le serveur Web.

Le programme script de ce formulaire récupère les paramètres entrés par l'utilisateur et les transmet à un programme en langage C.

Ce programme fait office de passerelle (gateway) entre la base Wais et le serveur Web. Il permet de visualiser la réponse à une requête directement depuis le serveur, c'est la raison pour laquelle les notices de la base sont écrites en langage HTML.

¹⁸ voir le serveur du NCSA à <http://hoohoo.ncsa.uiuc.edu/cgi>

V. MAINTENANCE ET PROMOTION

1. REGLES REDACTIONNELLES POUR LES PAGES HTML

Les documents HTML visualisés sous éditeur HTML sont difficiles à lire. Les balises et l'écriture complexe des caractères accentués augmentent le manque de lisibilité.

Or ces textes devront être relus lors des corrections et mises à jour.

Il convient donc de se fixer quelques règles de présentation. Elles n'auront aucune incidence sur la présentation du document par le client sous Web.

Les règles que nous nous sommes fixés sont les suivantes :

- saut de ligne entre deux paragraphes
- balises écrites en majuscules
- balises se succédant les une aux autres isolées sur une ligne

2. ORGANISATION DES FICHIERS

L'administrateur d'un serveur Web peut organiser comme il l'entend son serveur, contrairement à Gopher où il faut respecter une structure arborescente pour les documents.

Mais, de même que le choix des informations à transmettre est important, il est primordial d'organiser le plus logiquement possible les fichiers de stockage des pages HTML et de toutes celles qui y sont liées, notamment lorsque le serveur doit être déclaré au directory of servers. Quand il sera connu il va très vite être référencé par d'autres serveurs et le déplacement des fichiers sera, dès lors, déconseillé.

Dans notre cas, il s'agit surtout de faciliter la tâche au documentaliste chargé de la mise à jour des informations, puisque le serveur ne sera pas accessible de l'extérieur.

Ainsi, il est souhaitable de conserver une trace écrite de l'organisation thématique du serveur et des noms de fichiers des documents HTML.

Nous avons adopté l'organisation suivante : sur la machine serveur,

le répertoire **/Public/Mosaic/trillat/**

contient

/Pages : pages HTML simples (fichiers à extension html)

/Formulaires : pages de formulaires HTML (fichiers à extension html)

/Icons : icônes utilisés pour illustrer les pages du serveur (fichiers à extension gif)

/Wais-database : tous les fichiers concernant la base Wais: (waissource, les fichiers créés par la commande waisindex, le fichier de description de la base .fmt ...)

le répertoire **/Public/Mosaic/scripts/** (qui est le répertoire htbina) contient tous les programmes scripts, leur nom étant le même que celui des formulaires correspondants mais avec une extension sh.

Un serveur Web doit être régulièrement alimenté et "nettoyé" afin de toujours proposer une information fraîche et pertinente. Il ne faut donc pas négliger ou sous-estimer cette tâche qui peut sembler rébarbative mais n'est aucunement un travail de saisie classique.

En effet, elle implique des connaissances informatiques (essentiellement langage Unix) et documentaires. Le documentaliste chargé de la mise à jour doit connaître le fonctionnement des serveurs Web et leurs outils de conception, principalement les éditeurs HTML.

Il sera donc essentiel d'élaborer un guide pour la mise à jour et utile de réserver une demi-journée à donner quelques explications plus concrètes au futur gestionnaire.

3 AIDE A LA MISE A JOUR DE LA BASE WAIS

3.1...CONSIGNES DE REDACTION DES NOTES TECHNIQUES

Concernant la base Wais, la mise à jour, même si elle doit être confiée à une seule personne, en implique plusieurs : depuis les rédacteurs des notes techniques, à l'attention desquels sera élaboré un guide de rédaction (afin de faciliter la traduction en document HTML) jusqu'aux secrétaires qui sont chargés de récupérer le document original sous forme électronique et de le transmettre à la bibliothèque.

3.2...PROGRAMMES POUR SAISIR DES MODIFICATIONS

Afin de faciliter la mise à jour de la base Wais, trois programmes en C ont été élaborés.

En automatisant partiellement la saisie, ils diminuent les risques d'erreur et de disparités entre les notices et garantissent une validité permanente au fichier fmt de description de la base.

Ces programmes permettent de supprimer, ajouter ou modifier une notice.

3.2.1 remnotice

Il permet de supprimer une notice. Quand le numéro de notice a été entré, il affiche la notice correspondante et demande confirmation avant suppression. La liste numérotée des notices, accessible par une ancre depuis les formulaires d'interrogation, est alors automatiquement remise à jour.

Ce programme est disponible en annexe V, p.62.

3.2.2 addnotice et modnotice

Ces programmes commencent également par afficher la notice afin de vérifier qu'il s'agit de la bonne.

Ils fonctionnent ensuite sur des modes assez proches : ils permettent d'ajouter une notice (addnotice) ou de la modifier (modnotice).

Ils présentent chaque champ à remplir ou à modifier. Si le champ est obligatoire, le passage au champ suivant est impossible tant qu'aucune valeur n'a été saisie.

Les champs définis comme étant obligatoires sont les suivants : le nom du premier auteur, le titre, le premier lieu de rédaction et la référence (numéro séquentiel permettant d'identifier une note technique).

Un "format" de rédaction des notices a été défini : titre en majuscules, noms d'auteurs en majuscules suivi de l'initiale du prénom entre parenthèses... Au moment de toute saisie, il est rappelé en exemple.

A l'issue de toute modification de la base, le programme d'indexation waisindex est lancé automatiquement afin que les changements soient immédiatement pris en compte.

De nombreux tests ont été réalisés afin d'éprouver la fiabilité des programmes de mise à jour. Toutefois, cette fiabilité ne sera garantie que lorsque la base contiendra un nombre assez important de notices et qu'elle fonctionnera en situation réelle.

4. FAIRE CONNAITRE LE SERVEUR

Il est regrettable que le stagiaire ayant mis en place le serveur du CNS n'aie pas eu le temps de faire sa promotion, ce dont aurait vraisemblablement bénéficié notre serveur.

Différentes actions sont envisageables pour promouvoir le serveur et convaincre les chercheurs de son utilité. La bibliothèque, fréquentée par la plupart d'entre-eux, sera l'un des premiers lieux où ils auront connaissance de son existence et pourront le lister.

Concernant les personnes travaillant déjà au CNS, beaucoup utilisent une messagerie interne. Elle est également employée pour toutes les annonces officielles. Dans ce cadre, un message annonceur et la proposition de séances de démonstration pourrait inciter à tester ce nouveau mode d'information.

Spécialement pour les nouveaux arrivants, une ancre vers un guide imprimable a été mise en place. Ce document d'environ quatre pages est en fait un seul document HTML ce qui en facilite l'impression.

Une action peut également être menée à destination des thésards. En effet, une enquête a révélé qu'ils souhaiteraient que la bibliothèque leur soit systématiquement présentée lorsqu'ils arrivent au CNET. Lors de cette visite, le serveur et ses fonctionnalités seront décrits. Il leur sera utile notamment pour se familiariser avec le mode de fonctionnement de la bibliothèque, savoir quels titres de périodiques sont disponibles ou encore accéder depuis leur bureau aux notes techniques de la base Wais.

L'installation d'outils statistiques mesurant le taux de fréquentation de nos pages permettra d'observer le comportement des usagers et peut être quelles sont les pages les plus utilisées.

Concernant la base Wais, dans la commande de création du serveur, nous avons ajouté l'option-e `waisserver` qui permet d'alimenter un fichier qui tient un historique de toutes les interrogations. Ce fichier peut ensuite être exploité avec des commandes Unix pour élaborer des statistiques.

Enfin, une ancre vers la boîte aux lettres électronique de l'un des documentalistes offrira à l'utilisateur la possibilité de donner ses impressions et de faire part de ses commentaires qui pourront être pris en compte dans l'évolution des pages du serveur.

CONCLUSION

A ce jour, notre serveur n'est pas entièrement opérationnel. Le problème d'accès aux documents complets est en passe d'être résolu, deux types d'accès seront proposés selon que l'utilisateur travaille sur un poste Mac ou PC ou sur un poste Unix.

Sur PC et Mac, la lecture se fera directement sous Word 6, après enregistrement dans le répertoire courant, du fichier au format RTF. La version au format Postscript imprimable est également proposée. Concernant les machines Unix, aucune solution satisfaisante n'a été trouvée à ce jour.

Ce serveur ne concerne que la bibliothèque mais un autre projet, ayant trait à la circulation de l'information entre les CNET et à la veille technologique, est actuellement en marche. Ce projet intitulé RITS (Réseau d'Information technique et Stratégique) sera développé au travers d'un serveur Web et permettra, par le biais d'une base de données (Wais éventuellement), de proposer un accès à des documents stratégiques. Il présentera également une liste de personnes ressources chargées de guider les chercheurs et de récupérer les documentations sauvages stockées dans les bureaux.

Dans ce cadre, le serveur de la bibliothèque peut être considéré comme exemple intéressant du mode de fonctionnement d'un serveur Web et une maquette du type de développement réalisable au Cnet, notamment pour proposer l'accès à un document complet en ligne.

ANNEXES

Annexe I

FORMULAIRE SIMPLE D'INTERROGATION DE LA BASE WAIS



Pour en savoir plus...

Pour interroger la base, entrez le(s) terme(s) (*en minuscules et sans accents*) définissant votre recherche. Si vous interrogez pour la première fois, reportez vous à l'icône.

Auteur 1 : (Pour les noms composés, voir l'icône)
OU

Auteur 2 :
OU

Auteur 3 :

ET

Terme 1 :
OU

Terme 2 :
OU

OU

Terme 3 :

(Terme = mot du résumé ou mot du titre. N'entrez qu'un seul mot par case)

ET

Site de publication :

Ce formulaire est sommaire mais vous pouvez interroger sur un plus grand nombre de champs en utilisant l'autre formulaire, plus complexe.

Liste de l'ensemble des notices de la base.

Annexe II

FORMULAIRE DETAILLE D'INTERROGATION DE LA BASE WAIS



Pour en savoir plus...

Pour interroger la base, entrez le(s) terme(s) (*en minuscules et sans accents*) définissant votre recherche.

Un "ET" les relie entre eux.

Vous pouvez remplir au maximum 6 cases en même temps.

Si vous interrogez pour la première fois, reportez vous à l'icône.

Si vous préférez l'écran de recherche simple, vous pouvez y revenir.

Auteur(s) 1 :	<input type="text"/>	2 :	<input type="text"/>
(Pour les noms-composés voir l'icône ci-dessus)			
Mot du titre :	<input type="text"/>		
(un seul mot sans article)			
Descripteur(s) 1 :	<input type="text"/>	2 :	<input type="text"/>
(un seul par case, toujours au singulier)			
Date (ex: 1995) :	<input type="text"/>		
Type de document :	<input type="text" value="non_connu"/>		
Nom du responsable :	<input type="text"/>		
Mot(s) du resume 1 :	<input type="text"/>	2 :	<input type="text"/>
Site de publication :	<input type="text" value="non_connu"/>		
Groupement :	<input type="text"/>		

Liste de l'ensemble des notices de la base.



Retour au menu principal

Annexe III : FORMULAIRE DE SUGGESTION D'ACHAT

```

<TITLE>proposition d'acquisition</TITLE>
<CENTER>
<IMG SRC="/trillat/Icons/formulaire.gif">
<H3> FORMULAIRE DE SUGGESTION D'ACHAT
</CENTER>
<BR><BR>

<FORM ACTION="http://161.106.20.51/htbin/acquisitions.sh"
METHOD="GET">
<IMG SRC="/trillat/Icons/blueline.gif">
<BR><BR>
Merci de remplir un formulaire par document, le plus
précisément possible afin que votre demande soit
traitee dans les plus brefs delais :
<BR>

<PRE>
<HR>
Titre : <INPUT NAME="titre" SIZE=90>
Auteur (s) : <INPUT NAME="auteur" SIZE=90>
Editeur : <INPUT NAME="editeur" SIZE=40>
Date de publication : <INPUT NAME="annee" SIZE=40>
Numero de volume : <INPUT NAME="vol" SIZE=40>
<HR>
Nom du demandeur : <INPUT NAME="nom" SIZE=40>
Groupement/Departement : <INPUT NAME="group" SIZE=20>
Autres renseignements utiles: <INPUT NAME="mess1" SIZE=90>
(ISBN...) <INPUT NAME="mess2" SIZE=90>
<INPUT NAME="mess3" SIZE=90>
<INPUT NAME="mess4" SIZE=90>
<INPUT NAME="mess5" SIZE=90>
<INPUT TYPE="submit" VALUE="Envoyer la demande "> <INPUT
TYPE="reset" VALUE="Effacer les champs" >
</PRE>

</FORM>

<H3><A HREF="/trillat/Pages/presentation.html"><IMG
SRC="/trillat/Icons/back.gif">Retour au menu principal </A></H3>

```

SCRIPT DE TRAITEMENT

```
#!/bin/sh
FILE1="/export/user/Public/Mosaic/NOBODY/suggestions.xls"
export QUERY_STRING
eval `/export/user/Public/Mosaic/scripts/cgiparse -form`
echo Content-Type: text/html
echo
echo '<PRE><H4>'

FORM_titre=${FORM_titre:=0}
FORM_auteur=${FORM_auteur:=0}
FORM_editeur=${FORM_editeur:=0}
FORM_annee=${FORM_annee:=0}
FORM_vol=${FORM_vol:=0}
FORM_nom=${FORM_nom:=0}
FORM_group=${FORM_group:=0}
FORM_mess1=${FORM_mess1:=0}
FORM_mess2=${FORM_mess2:=0}
FORM_mess3=${FORM_mess3:=0}
FORM_mess4=${FORM_mess4:=0}
FORM_mess5=${FORM_mess5:=0}

NUM=`/usr/bin/tail -1 $FILE1 | awk '{print $1}'`
NUM=`expr $NUM + 1`
D=`/usr/bin/date '+%d-%b-%Y'`
/usr/bin/echo
"$NUM\t$D\t$FORM_titre\t$FORM_auteur\t$FORM_editeur\t$FORM_annee\t$FORM_vol\t$FORM_nom\t$FORM_group\t$FORM_mess" >>$FILE2

echo "Votre requete a ete soumise."
echo '</H4></PRE>'
```

Annexe IV : FICHER FMT DE DESCRIPTION DE LA BASE

```
<record-end> / ^ < \ notice > /  
< layout >  
< headline > / ^ < titre > / < \ titre > / 80 / ^ < titre > /  
< end >
```

```
< field > / ^ < auteur > /  
TEXT GLOBAL  
< end > < \ auteur > /
```

```
< field > / ^ < titre > /  
TEXT GLOBAL  
< end > / ^ < \ titre > /
```

```
< field > / ^ < auteur > /  
au TEXT LOCAL  
< end > < \ auteur > /
```

```
< field > / ^ < titre > /  
ti TEXT LOCAL  
< end > < \ titre > /
```

```
< field > / ^ < descripteur > /  
de TEXT LOCAL  
< end > < \ descripteur > /
```

```
< field > / ^ < annee > /  
an TEXT LOCAL  
< end > < \ annee > /
```

```
< field > / ^ < reference > /  
ref TEXT LOCAL  
< end > < \ reference > /
```

```
< field > / ^ < type > /  
ty TEXT LOCAL  
< end > < \ type > /
```

```
< field > / ^ < responsable > /  
res TEXT LOCAL  
< end > < \ responsable > /
```

```
< field > / ^ < mot > /  
mot TEXT LOCAL  
< end > < \ mot > /
```

```
<field> /<site>/  
si TEXT LOCAL  
<end> /<Vsite>/
```

```
<field> /<groupement>/  
gr TEXT LOCAL  
<end> /<Vgroupement>/
```

EXEMPLE DE NOTICES EXTRAITES DE LA BASE

```

<notice>
<auteur>ALAIMO (M.)</auteur><P>
<auteur>LECLERCQ (J.L.)</auteur><P>
<titre>PROJET AXEL - RETICULE COMMUN - CNET/CENTRE COMMUN -
PRESENTATION DU JEU DE MASQUES - LES MODULES DE TEST</titre><P>
<mot>AXEL est un jeu de deux ou trois reticules permettant la realisation
de</mot>
<mot>capacites au Centre Commun et au CNET Meylan. Il repond aux
besoins en</mot>
<mot>fabrication de capacites mais egalement permet d'autres
etudes</mot>
<mot>neecessitant un ou deux niveaux (defauts induits par la gravure
du</mot>
<mot>plasma, motifs electriques pour lithographie, etude TiSi2
selectif,</mot>
<mot>BICMOS, ...). Apres une presentation generale du masque, nous
decrivons</mot>
<mot>(origine, objectif poursuivi, brochage du dessin) les
differeents</mot>
<mot>modules de tests electriques.</mot><P>
<annee>1995</annee><P>
<reference>NT/CNS/DCF/73</reference><P>
<site>Grenoble</site><P>
<type>note</type><P>
<groupement>DCF</groupement><P>
<doelectr><A HREF="ftp://user/limoux/Public/Mosaic/ftp-
anonyme/pub/Doc/catalogue.rtf">recuperation du fichier contenant le
texte integral au format RTF lisible par Word (Mac ou
PC)</A></doelectr><p>
<doelectr><A HREF="ftp://user/limoux/Public/Mosaic/ftp-
anonyme/pub/Doc/catalogue.ps">recuperation du fichier contenant le
texte integral au format Postscript imprimable
directement</A></doelectr><p>
</notice>
<hr>
<notice>
<auteur>TOURET (O.)</auteur><P>
<titre>MODELISATION ET SIMULATION DU DEPOT SELECTIF DE TiSi2 PAR
CVD</titre><P>
<annee>1995</annee><P>
<responsable>PONCET (A.)</responsable><P>
<reference>RS/CNS/DCF/95-23</reference><P>
<site>Grenoble</site><P>
<type>rapport</type><P>

```

```
<groupement>DFC</groupement><P>
<doelectr><A HREF="ftp://user/limoux/Public/Mosaic/ftp-
anonyme/pub/Doc/catalogue.rtf">recuperation du fichier contenant le
texte integral au format RTF lisible par Word (Mac ou
PC)</A></doelectr><p>
<doelectr><A HREF="ftp://user/limoux/Public/Mosaic/ftp-
anonyme/pub/Doc/catalogue.ps">recuperation du fichier contenant le
texte integral au format Postscript imprimable
directement</A></doelectr><p>
</notice>
<hr>
```

Annexe IV : PROGRAMME DE DESTRUCTION DE NOTICES

```

#include <stdio.h>

#define Max      256

int lirechaine ( char Chaîne [ Max ] ) ;
void liretexte ( char T [ Max ] ) ;
void cleartexte ( char T [ Max ] ) ;
void listnotice ( int N_notice ) ;

FILE *FPIN, *FPOUT ;

void main ( int argc, char *argv [ ] )
{
int Notice_courante = 0, Quelle_notice, COMPI ;
char Chaîne [ Max ], octet, Batch = 'I' ;
char File1 [ Max ], File2 [ Max ], Reponse [ Max ] ;

    File1 [ 0 ] = '\0' ;
    strcat ( File1, "/export/user/Public/Mosaic/trillat/wais-database/B1.html", Max
    File2 [ 0 ] = '\0' ;
    strcat ( File2, "/export/user/Public/Mosaic/trillat/wais-database/B1.html.O", Max
    Chaîne [ 0 ] = '\0' ;

/* Sauvegarde de la base */
/* ===== */

    FPIN = fopen ( File1, "r" ) ;
    FPOUT = fopen ( File2, "w" ) ;

    while ( ( octet = getc ( FPIN ) ) != EOF )
        putc ( octet, FPOUT ) ;

    fflush ( FPOUT ) ;
    fclose ( FPIN ) ;
    fclose ( FPOUT ) ;

/* Fin sauvegarde de la base */
/* ===== */

/* Nous voulons impérativement un numero de notice */
/* ===== */

    switch ( argc ) (
case 1 :      printf ( "Utilisation : %s Numero_de_notice\n", argv [ 0 ] ) ;
              exit ( 1 ) ;
              break ;

case 2 :      Quelle_notice = atoi ( argv [ 1 ] ) ;
              if ( Quelle_notice < 1 ) {
                  printf ( "La notice %d n'existe pas.\n", Quelle_notice
                  exit ( 1 ) ;
                  break ;
              }
              break ;

case 3 :      Quelle_notice = atoi ( argv [ 1 ] ) ;
              if ( Quelle_notice < 1 ) {
                  printf ( "La notice %d n'existe pas.\n", Quelle_notice
                  exit ( 1 ) ;
                  break ;
              }
              if ( strcmp ( argv [ 2 ], "B" ) == 0 ) Batch = 'B' ;
              break ;
    )

```

```

default :      printf ( "La notice %d n\'existe pas.\n", Quelle_notice ) ;
              exit ( 1 ) ;
              break ;
}

/* Ouverture de la base de sauvegarde */
/* ===== */

EPIN = fopen ( File2, "r" ) ;

/* Si nous sommes en interactif, est-ce la bonne notice ? */
/* ===== */

if ( Batch == 'I' ) {
    printf ( "Voici la notice que vous voulez detruire :\n\n" ) ;
    listnotice ( Quelle_notice ) ;

    printf ( "\n\nEst-ce la bonne notice (oui ou non) ? " ) ;
    cleartexte ( Reponse ) ;
    liretexte ( Reponse ) ;
    if ( strcmp ( Reponse, "oui" ) != 0 ) {
        fclose ( FPIN ) ;
        exit ( 0 ) ;
    }

    fclose ( FPIN ) ;
    FPIN = fopen ( File2, "r" ) ;
}

/* Destruction de la notice */
/* ===== */

FPOUT = fopen ( File1, "w" ) ;

while ( lirechaine ( Chaine ) != EOF ) {
    COMP1 = strcmp ( Chaine, "<notice>" ) ;
    switch ( COMP1 ) {

        case 0 : Notice_courante = Notice_courante + 1 ;
                if ( Notice_courante != Quelle_notice )
                    fprintf ( FPOUT, "%s\n", Chaine ) ;
                if ( Notice_courante == Quelle_notice ) {
                    lirechaine ( Chaine ) ;
                    while ( strcmp(Chaine, "<hr>") != 0 )
                        lirechaine ( Chaine ) ;
                    Notice_courante = Notice_courante + 1 ;
                }
                break ;

        default : fprintf ( FPOUT, "%s\n", Chaine ) ;
                break ;
    }
}

fclose ( FPIN ) ;
fclose ( FPOUT ) ;

system ( "/usr/local/wais/bin/waisindex -d B1 -export -T HTML -t fields /export/" )
}

int lirechaine ( char C [ Max ] )
{

```

```

int i = 0 ;
char O ;
while ( ( C [ i ] = fgetc (FPIN) ) != EOF ) {
    if ( C [ i ] == '\012' ) {
        C [ i ] = '\0' ;
        return ( i ) ;
    } else i = i + 1 ;
}
C [ i ] = '\0' ;
return ( EOF ) ;
}

void listnotice ( int N_notice )
{
char Ligne [ Max ] ;
int N_courant = 0, COMP ;

while ( lirechaine ( Ligne ) != EOF ) {
COMP = strcmp ( Ligne, "<notice>" ) ;
switch ( COMP ) {

case 0 :          N_courant = N_courant + 1 ;
                  if ( N_courant == N_notice ) {
                      printf ( "%s\n", Ligne ) ;
                      lirechaine ( Ligne ) ;
                      while ( strcmp(Ligne,"<hr>") != 0 ) {
                          printf ( "%s\n", Ligne ) ;
                          lirechaine ( Ligne ) ;
                      }
                      printf ( "%s\n", Ligne ) ;
                      N_courant = N_courant + 1 ;
                  }
                  break ;
default :        break ;
}
}
}

void liretexte ( char T [ Max ] )
{
int i = 0 ;

while ( ( T [ i ] = getchar ( ) ) != '\n' ) i++ ;
for ( i = 0 ; i < Max ; i++ ) if ( T [ i ] == '\n' ) T [ i ] = '\0' ;
}

void cleartexte ( char T [ Max ] )
{
int i ;
for ( i = 0 ; i < Max ; i++ )
    T [ i ] = '\0' ;
}

```

BIBLIOGRAPHIE

BIBLIOGRAPHIE

OUVRAGES :

L'Internet professionnel : témoignages, expériences, conseils pratiques de la communauté enseignement & recherche. Nancy : CNRS Editions, 1995. Collection CNRS et Université. 448p.
ISBN 2-271-05256-

Le traitement électronique du document. Cours INRIA, 3-7 octobre 1994, 1994, Aix-en-Provence. Ouvrage coordonné par Jean-Claude Le Moal et Bernard Idoine. Paris : ADBS Editions, 1994. Collection Sciences de l'information. 285p.
ISBN 1160-2376

ANDRIEU, Olivier. **L'Officiel 95 d'Internet.** Paris : Eyrolles, 1995. 253p.
ISBN 2-212-08904-X

KROL, Ed. **Le Monde Internet** : guide & ressources. 2ème édition. S.L. : Editions O'Reilly International Thomson, 1994. 543p.¹⁹
ISBN 2-84177-000-1

La QUEY, Tracy. **Sésame pour Internet** : initiation au réseau planétaire. France : Editions Addison-Wesley, 1994. 242p.
ISBN 2-87908-083-5

LEROY, Alain. **Le Système unix.** Télécom Bretagne : Département Informatique et Réseaux, Edition avril 1995. 70p.

NIEUTIN, Anne. **Un serveur Web à la Bibliothèque Nationale de France** : conception et réalisation. Mémoire de DESS Applications à la Télématique. Paris : Université Denis Diderot, 1995. 146p.

PELIKS, Gérard. **Le World Wide Web** : création de serveurs sur Internet. France : Editions Addison-Wesley, 1995. 220p.
ISBN 2-87908-102-5

¹⁹ mis à jour en ligne par l'éditeur à l'adresse suivante
<http://nearnet.gnn.com/wic/newrescat.toc.html>

SIBERTIN-BLANC, Martine. **Nouvelles technologies et communication de l'information** : des besoins des utilisateurs à l'ingénierie documentaire. Paris : ADBS Editions, 1994. Collection Sciences de l'information. 277p.

ISBN 2-901046-71-1

THOMAS, Brian J. **The Internet for scientists and engineers** : Online Tools and Resources. Washington (USA) : SPIE, 1995. VIII-450 p.

ISBN 0-8194 -1806 -4

TITTLE, Ed & STEVE, James. **HTML for dummies**. Forest City (CA) : IDG Books, 1995. XVIII-452p.

ISBN 1-56884-330-5

ARTICLES :

Internet Technologies : Today and Tomorrow. Solutions showcase Demonstration. { Brochure édité conjointement par UB Networks et Bull pour les démonstrations aux journées Networld+Interop 95 } . Paris : Parc des expositions de Paris, 11-15 Septembre 1995. { 6 pages }.

Ghesquiere, Gilles. **Qui a peur d'Internet?** Solutions télématiques, N°35, mai 1995, p.16-17.

Levy-Abegnoli, Thierry. **Le World Wide Web tisse sa toile sur la planète.** 01 Informatique, N°1359, 26 mai 1995, p.18.

Levy-Abegnoli, Thierry. **Serveurs WEB : les premiers outils de développement arrivent.** 01 Informatique, N°1355, 28 avril 1995, p.22.

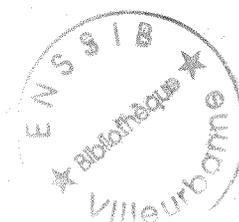
Bodan, Fabienne. **Dossier Internet.** Solutions télématiques, N°37, juillet-août 1995, p.30-33.

Bodan, Fabienne. **Le coût de développement d'un serveur Web.** Solutions télématiques, N°35, mai 1995, p.35.

Devalal, Danièle. **Autoroutes de l'information : un nouveau schéma juridiques des réseaux.** Solutions télématiques, N°35, mai 1995, p.46.

Du Castel, François. **Autoroutes de l'information et société de communication.** Réseaux : Communication technologie société, N°71, mai-juin 1995, p.109-114.

ISSN 075-7971



QUELQUES SERVEURS INTERESSANTS

(Les adresses ci dessous sont différentes de celles indiquées au fil des pages)

Documentation sur Netscape

<http://home.mcom.com/home/welcome.html>

Généralités sur WWW

<http://www.w3.org/hypertext/WWW/TheProject.html>

<http://urec.fr/www/docs/WWW/>

http://www.hcc.hawaii.edu/guide/www_guide.html

Généralités sur HTML

<http://www.w3.org/hypertext/WWW/MarkUp/MarkUp.html>

<http://www.urec.fr/docs/WWW/WWW.html>

Logiciels relatifs à HTML

<http://www.urec.fr/docs/www-soft.html>

<http://info.cern.ch/hypertext/WWW/Daemon/Overview.html>

Le langage HTML

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

<http://www.urec.fr/docs/WWW/>

[puis chapter2-3.html](#)

Guide de style du document HTML

<http://www.willamette.edu/html-composition/strict-html.html>

Guide de l'utilisateur de RTFtoHTML

<ftp://ftp.primaie.uiuc.edu/pub/RTF>

Généralités sur Wais

<ftp://urec.fr>

dans pub/reseaux/services-infos/wais/docs_fr/WAIS.Barthelemy.txt

<ftp://zenon.inria.fr> **dans** <pub/wais>

<ftp://cirm.univ-mrs.fr> **dans** <pub/wais/wais1>

LISTES DE DISCUSSION

wais-fr@univ-lyon1.fr : liste intéressant les administrateurs de bases Wais

www.fr@univ-rennes.fr : liste traitant des serveurs WWW

biblio.fr@univ-rennes.fr : liste destinée aux membres de l'ADBS

BIBLIOTHEQUE DE L'ENSSIB



966547C