

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

1000

**UNIVERSITE**  
**CLAUDE BERNARD**  
**LYON I**

## **DESS en INFORMATIQUE DOCUMENTAIRE**

### **Rapport de Stage**

**Projet de développement du Système d'Information de la Mission  
Handicap , Université Claude Bernard Lyon I.**

María Rebeca Bilbao Acuña

Sous la direction de  
Sylvie Laine-Cruzel

ENSSIB et Université Claude Bernard Lyon I

**1995**

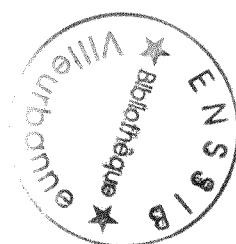
22

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

**UNIVERSITE**  
**CLAUDE BERNARD**  
**LYON I**

## **DESS en INFORMATIQUE DOCUMENTAIRE**

**Rapport de Stage**



**Projet de développement du Système d'Information de la Mission  
Handicap , Université Claude Bernard Lyon I.**

María Rebeca Bilbao Acuña

Sous la direction de  
Sylvie Laine-Cruzel

ENSSIB et Université Claude Bernard Lyon I

**1995**

1995

EST

22

## **Remerciements**

Je tiens à remercier toute l'équipe de la Mission Handicap, qui, par la gentillesse et ses précieux conseils, m'ont aide et encouragé tout au long de ce stage. Et ont ainsi contribué à rendre cette expérience personnelle et professionnelle aussi intéressante qu'enrichissante.

**Titre:** Projet de développement du Système d'Information de la Mission Handicap, Université Claude Bernard - Lyon I.

**Auteur:** María Rebeca Bilbao Acuña

**RESUME :** Etude du Serveur Handy W W W de la Mission Handicap , Université Claude Bernard - Lyon I, concernant l'analyse de l'existant et l'évaluation des services pour l'amélioration de ceux-ci, à travers la recherche d'un convertisseur de fichiers traitement texte, RTF à fichiers HTML ; la création d'une base de données Access pour statistiques et la création à niveau de maquette d'un catalogue en ligne WAIS.

**DESCRIPTEURS :** SERVEUR W W W / HANDICAPES / CONVERTISSEURS / FICHIER WORD / FICHIER RTF / FICHIER HTML / STATISTIQUES / CATALOGUE / CAHIER DES CHARGES /

**ABSTRACT :** Study of the Server Handy WWW of Handicap Mission, University Claude Bernard Lyon I , concerning the analysis of the present state and the evaluation of services in order to improve them thanks : the research of a filter of files WORD, RTF , HTML; the creation of a database ACCESS for the statistics and the creation of a catalog on-line WAIS.

**KEYWORDS:** SERVER WWW / HANDICAPED STUDENTS / FILTERS / FILES WORD / FILES RTF / FILES HTML / STATISTICS / CATALOG / SPECIFICATIONS /

# **TABLE DE MATIERES**

## **INTRODUCTION**

### **I.- ANALYSE DE L'EXISTANT**

1.- Demandes de la Mission.....	1
---------------------------------	---

### **II.- DESCRIPTION DE LA MISSION HANDICAP**

2.1.-Présentation.....	2
2.1.1.- Comment est née la Mission Handicap.....	2
2.1.2.- L'aide aux étudiants handicapés.....	3
2.1.3.- Recherche et développement.....	5

### **III.- LE SERVEUR DE LA MISSION HANDICAP**

3.1.- Introduction.....	6
3.2.- Le serveur Handy sur WWW.....	6
3.3.- Description et structure du serveur Handy.....	7
3.3.1.- Objectif.....	7
3.3.2.- Description.....	7

### **IV.- UTILISATEURS DU SERVEUR HANDY**

4.1.- Usagers.....	12
4.2.- Les formations à l'Université.....	13

### **V.- PROCEDURE DE COLLECTE D'INFORMATION ET TRAITEMENT DES DOCUMENTS ELECTRONIQUES**

5.1.- Gestion électronique de documents.....	15
a) Qu'est-ce qu'un document.....	15
b) Les usages de documents.....	16
c) La collecte d'information.....	16
d) La chaîne documentaire.....	17
e) Récupération et structuration HTML.....	19

## **VI.- BALISAGE HTML**

6.1.- Introduction à SGML.....	20
6.2.- Utilisation de HTML à la Mission Handicap.....	21
6.3.- Indexation et recherche de documents.....	23
6.4.- Recherche de documents.....	23

## **VII.- ETUDE DES BESOINS**

7.1.- Introduction.....	23
7.2.- Objectifs.....	24
7.3.- Analyse des applications à réaliser.....	25

## **VIII.- ETUDE RELATIF A L'UTILISATION DU SERVEUR**

8.1.- Statistiques.....	26
8.1.1.- WUsage 3.2.....	26
8.1.2.- Getstats 1.2.....	27
8.1.3.- WWW Stats 1.0.....	28
8.1.4.- Base de données Access.....	28
8.2.- Enquête.....	30

## **IX.- ETUDE CONCERNANT AU BALISAGE HTML**

9.1.- Dysfonctionnement et nature du problème.....	32
9.2.- Tâches réalisées.....	33
9.2.1.- Dlimitation des besoins et recherche en Internet.....	33
9.2.2.- Etude des résultats obtenus.....	33
9.2.3.- Sélection de convertisseurs.....	33
9.2.4.- Essai du convertisseur choisi.....	36

## **X.- CREATION D'UN CATALOGUE EN LIGNE**

10.1.- Etude du Logiciel Doris/Loris.....	36
10.2.- Etude de la collection existante.....	37
10.3.- Etude des données bibliographiques.....	38

10.4.- Visulisations des futurs problèmes de recherche.....	41
10.5.- Maquette de la fiche bibliographique.....	43
10.6.- Procèdures.....	44

## **CONCLUSION**

## **ANNEXES**

### **PROJETS DE LA MISSION**

- Projet Prince
- Projet Cristallin
- Projet Capesa

### **STATISTIQUES**

- WUsage 3.2
- Getstats Version 1.2
- WWW Stats
- Base de Données Access
  - Tables
  - Relations de la Base de données
  - Requêtes définies
  - Graphiques et tableaux

### **CONVERTISSEURS**

- Cyberleaf
- Rtfhtml
- Megadots
- Dotsplus

## INTRODUCTION

Le présent document est le résultat du stage réalisé à la Mission Handicap de l'Université Claude Bernard- Lyon I. Cette activité académique essentielle pour obtenir le diplôme de DESS en Informatique Documentaire a eu une durée de quatre mois et son objectif fondamental a constitué la rédaction d'un Cahier des Charges pour la Mission.

Au début du stage , les besoins de la Mission ont été la création d'un catalogue en ligne (sous DORIS/LORIS) pour cataloguer la collection des documents administratifs, scientifiques et techniques existants au sein de l'entité. Après la réalisation d'une première étude préliminaire de l'existant et d'une série des réunions entre le responsable de la Mission, la directrice du stage à l'Enssib et la stagiaire, on a décidé de définir l'objectif du stage et les tâches à réaliser pendant la période de quatre mois.

Les demandes finales de la Mission ont été:

1.- Etude de l'existant et la rédaction d'un Cahier des Charges concernant au projet de création du catalogue en ligne et la définition d'un scénario capable d'inclure toutes les possibles applications informatiques à réaliser a court terme.

2.- Développement du Cahier des Charges et l'élaboration d'une maquette basée sur:

a) L'utilisation du logiciel DORIS/LORIS pour la gestion du catalogue de documents électroniques.

b) Etude concernant à la collecte d'information des documents scientifiques et techniques faits par les enseignants de l'Université.



c) Etude concernant au balisage HTML pour chaque type de document au sein de la Mission afin d'améliorer leurs procédures.

Il faut détacher qui pendant la réalisation du stage, on a évalué le scénario et alternatives proposées par la stagiaire afin de délimiter les priorités à suivre.

Dans ce processus on s'aperçut qu'existe une problématique de définition des objectifs au sein de la Mission et ce par celui-ci que pour rédiger le Cahier des Charges correspondant il faut qu'on passe par une étude de l'existant très approfondi, afin de connaître les objectifs de l'entité, leurs utilisateurs et ses possibilités réelles de développement pour améliorer les services en ligne sur le serveur Handy.

Les étapes réalisées sont:

- a) Etude de l'existant.
- b) Présentation du scénario.
- c) Evaluation du serveur et services proposés aux utilisateurs.
- d) Réalisation des statistiques sous différents logiciels pertinents afin de connaître le niveau d'utilisation réel du serveur.
- e) Etude de la procédure de collecte et traitement des documents électroniques.
- f) Définition des dysfonctionnements du système.
- g) Recherche en ligne et contacte avec spécialistes au étrangère, avec l'objectif de trouver des logiciels et des convertisseurs pour améliorer la tâche de conversion des fichiers traitement texte, RTF à fichiers HTML.
- h) Etude de l'ensemble DORIS/LORIS et sa possible application à la Mission.

i) Etude et présentation des alternatives pour la création du catalogue en ligne et la liaison au document hypertexte.

j) Réalisation du catalogue à niveau de maquette.

Un point relevant de cette activité à été de donner à la Mission des recommandations pour améliorer son système d'information sur la base d'une vision comme bibliothécaire, documentaliste et étudiante d'informatique documentaire.

**ENTITE : MISSION HANDICAP - UNIVERSITE CLAUDE BERNARD - LYON I.**

**RESPONSABLE : Mr. CLAUDE DECORET**

## **I.- ANALYSE DE L'EXISTANT**

### **1.1.- DEMANDES DE LA MISSION::**

Rédaction d'un cahier des charges concernât le projet à réaliser. Il faudra trouver une méthodologie d'introduction d'informations documentaires sur des supports électroniques accessibles en réseau.

Au texte intégral, il faut associer la définition d'une structure des données à usage documentaire. Le texte intégral doit être géré dans un format reexploitable et facilitant son introduction dans le serveur Handinet. Cette partie s'étend sur une période du début de stage jusqu' au Juillet.

La deuxième partie qui s'achève en fin septembre, concerne le développement du cahier des charges et l'élaboration d'un prototype en basé sur:

- a) L' utilisation de DORIS/LORIS pour la gestion des données bibliographiques.
- b) Une collecte de documents pertinents écrits en français.
- c) Une étude concernant le travail à faire pour chaque type de documents et chaque origine, afin d'étudier les solutions possibles pour le balisage HTML.

## **II .- DESCRIPTION DE LA MISSION HANDICAP**

### **2.1.- PRESENTATION :**

**UNIVERSITE CLAUDE BERNARD LYON I**

**MISSION HANDICAP**

Bâtiment 702, 43 Bd du 11 Novembre 1918  
F-69622 VILLEURBANNE CEDEX - FRANCE  
tél.: (+33) 72 44 82 79 ou  
(+33) 72 43 14 77 fax: (+33) 72 43 15 05  
e-mail: [decoret@handy.univ-lyon1.fr](mailto:decoret@handy.univ-lyon1.fr)  
Responsable: Claude DECORET

#### **2.1.1.- Comment est née la Mission Handicap?**

##### **L'idée:**

Pour mieux comprendre le rôle et l'importance de la Mission handicap, il faut remonter 10 ans en arrière. En effet, c'est au cours de l'exposition pour les 30 ans du Comité Louis Braille (association que regroupe des associations de non-voyants à Lyon) que germe l'idée que l'informatique est de plus en plus importante pour les déficients visuels. Suit alors une demande de subvention, qui donne lieu en 1985 à un premier crédit et également au détachement d'une personne non voyante de Paris, M Jalade (inspecteur de finances).

Ainsi, à Lyon des handicapés visuels commencent à suivre des cours de remise à niveau en informatique.

##### **Par la suite...:**

Entre temps, en 1983 et 1984, la FIDEV (Formation Informatique pour les Déficiants visuels) a été constituée. En 1985, la première circulaire qui organise le Tiers temps des examens est diffusée. Ce système permet aux étudiants handicapés de bénéficier d'un tiers du temps réglementaire en plus; ce temps étant nécessaire au bon déroulement de l'examen.

Puis, le 18 Janvier 1990, une journée nationale de sensibilisation et d'échanges sur l'accueil des étudiants handicapés a lieu à l'Université Claude Bernard Lyon I. Elle regroupe environ 70 personnes qui tentent dans différentes universités de mettre sur pied un service d'accueil aux étudiants handicapés.

### **La naissance et les premiers pas:**

En 1991, le ministère demande à chaque université d'avoir une personne chargée de mission pour les handicapés. Ainsi naît la Mission Handicap, et on trouve alors une personne, ou même parfois un service, dans chaque université, chargé de s'occuper de tout ce qui concerne les étudiantes handicapées.

C'est ainsi que Mr. Decoret, directeur de recherche, est mis à disposition par le CNRS, pour tenir ce poste à l'université Lyon I, à mi-temps.

Viennent ensuite en 1992, les premiers crédits pour Lyon I; ils servent à équiper la Bibliothèque Universitaire en ordinateurs et autres matériels informatiques. En 1993, des crédits de fonctionnement sont accordés en plus des crédits d'équipement. Il ont permis de mettre en place un service de répétition de cours et d'assistance aux examens.

### **Actuellement**

Depuis le 1er avril 1994, M. Decoret s'occupe à plein temps de la Mission Handicap à Lyon I. Depuis la rentrée 94, la Mission Handicap est devenue un service à part entière. Celui-ci se compose d'un chargé de mission rattaché à la présidence de l'Université Lyon I: M. Claude Decoret directeur de recherche au CNRS; de deux ingénieurs informatiques. M. Laurent De Berti Service Civil National et M. François Bayard Ingénieur d'études C.N.R.S. (à 20%), un secrétaire administratif: M. Georges Brun; un C.E.S.: Mlle Lydie Morin.

## **2.1.2.- L'aide aux étudiants handicapés:**

### **Généralités:**

L'accès aux études universitaires par les personnes handicapées est facilité par l'utilisation d'ordinateurs. De plus en plus de matériel se présente sous forme électronique et ce matériel semble être également de plus en plus accessible pour les handicapés. Ainsi, le nombre d'étudiants handicapés augmente petit à petit et de nombreuses universités ont créé des services pour faciliter leurs études.

En France, chaque université (soit environ 90) a un représentant chargé de mission pour les étudiants handicapés et de plus, le Ministère a un service spécial de coordination de toutes leurs actions (Ministère de l'Enseignement Supérieur et de la Recherche).

Dans la région Rhône-Alpes, 3 universités ont développé des services, chacune avec une orientation spécifique.

## A Lyon 1:

Il existe une synergie entre plusieurs services qui sont:

- la Commission Handicap
- le service médical pour les étudiants,
- la Mission Handicap
- les professeurs.

Définissons le rôle de chacun de ces services par rapport aux étudiants handicapés

**a) La Commission Handicap:** par définition , elle est composée d'étudiants handicapés. Il y a environ 40 étudiants handicapés pour un total de 24 000 étudiants à l'université Lyon I.

**b) Le service de prévention médicale:** Ce service inter universitaire est le seul qui soit qualifié pour donner le droit à un étudiant de passer son examen avec des dispositions spéciales (ces dispositions proviennent de la circulaire Ministère sur le Tiers temps). L'étudiant, disposera de l'aide d'une personne qualifiée, ou encore de la possibilité de passer ses examens dans une pièce séparée.

**c) La Mission Handicap:** elle s'occupe des dispositions à prendre pour que les étudiants puissent passer leurs examens et effectuer leur études dans les meilleures conditions possibles. De plus, la Mission Handicap, en collaboration avec la Commission Handicap, travaille pour faire connaître son existence à tous les étudiants.

### Les moyens disponibles:

La Bibliothèque Universitaire, en collaboration avec la Mission Handicap, dispose d'une pièce équipée de différents machines, cette pièce est accessible à certaines heures par les étudiants handicapés. De plus, une personne est présente 2 heures par jour pour éventuellement aider les étudiants à utiliser les appareils.

Les équipements sont plutôt orientés pour les déficients visuels, c'est à dire: deux ordinateurs avec une synthèse vocale (anglaise et française) Apollo, une imprimante Braille INDEX, un lecteur d'écran "Screen Reader" (IBM), une imprimante normale et une machine à lire "READIND EDGE".

Les étudiants ont également la possibilité de se connecter au réseau et de découvrir le monde d' INTERNET. Ils ont quelques heures de cours pour se familiariser avec le réseau et obtiennent ensuite une adresse électronique et un compte.

## **La station de travail à la Mission handicap:**

La mission dispose d'une station de travail Sun Microsystems, SPARC 20, 32 Mo de RAM, Bioprocresseurs, 2 Go de disque dur, ectan 20 pouces.

### **2.1.3.- Recherche et développement:**

#### **Le rôle de la Mission Handicap**

Le rôle d'une Mission Handicap étant de faciliter l'accès aux études pour les étudiants handicapés, implique donc aussi la mise en oeuvre de divers matériels, logiciels, etc. Si l'on tient compte de l'évolution rapide des techniques, la Mission Handicap doit se tenir informée de toute innovation susceptible d'aider les étudiants handicapés dans leurs études universitaires. Il est aussi souhaitable que la Mission Handicap puisse également proposer des recherches et développements reflétant les besoins des utilisateur. Ce rôle a été facilité dans notre université du fait de ses capacités de recherches dans le domaine scientifique et de l'ancienne carrière scientifique de l'actuel responsable du service. Les besoins qui sont apparus concernent deux domaines: la transcription en braille des textes mathématiques et les interfaces homme-machine dans l'accès à l'information électronique.

#### **Transcription braille des textes mathématiques**

Il s'agit d'une transcription automatique de textes écrits en TeX ou LaTeX en braille, soit sur papier à partir d'embousseuses liées à l'ordinateur, soit sur les pages braille éphémères des terminaux spécialisés courants. Cette recherche a été effectuée par un stagiaire de l' IUT d'informatique auprès de la Mission handicap, ce stagiaire ayant par la suite continué ce travail en effectuant son service militaire comme service civil auprès de la Mission Handicap. Il s'agit d'un développement de haute technicité tant du point de vue analyse que du point de vue informatique (connaissance de Tex, du braille mathématique, de LEX et YACC, etc.).

#### **Interface homme-machine**

Il s'agit dans le cadre du développement de stations ou poste de travail multimodal et multimédia adaptés pour non-voyants entrepris à l' **IRIT** à Toulouse (**UPS**) d'obtenir un accès aux textes scientifiques et techniques par le marquage SGML des expression mathématiques afin de les rendre accessibles aux synthèses vocales. Cette étude se situe dans le cadre d'une thèse de l' **IRIT**.

**La Mission Handicap** participa aux spécifications nécessaires à l'utilisation de **PRINCE** (Voir annexe PROJETS DE LA MISSION HANDICAP), dans le cadre du D.E.S.S. Direction des Etablissements Sanitaires et sociaux à Lyon 1 en cours d'habilitation et plus généralement à l'utilisation par les étudiants et enseignants de cette Université.

Outre la participation à l'analyse des besoins et à la validation de **CRISTALLIN**, le programme de l' **U..C.B.L.** s'articule autour de la récupération et la structuration SGML de documents.

### III.- LE SERVEUR DE LA MISSION HANDICAP:

#### 3.1.- Introduction:

Le serveur de la Mission Handicap est lié aux différents serveurs de l'Université Claude Bernard. Il permet l'accès à toutes les informations disponibles sur le campus et aussi de se connecter à d'autres serveurs WWW de l' UCBL ou du monde entier. Il inclut une série de services en ligne dirigés vers les élèves handicapés et tout l'univers des élèves de l' UCBL.

#### 3.2.- Le serveur Handy sur WWW.

##### a) World - Wide Web:

The World Wide Web a été développée en Genève, Suisse dans le European Center of Nuclear Research. Son catalogue de ressources est rangé par sujet et chaque source se base en hypertexte. L'information en hypertexte peut faire la liaison des différents documents électroniques. HTML est le langage qui permet de créer des documents en W3, et l'URL ( Uniform Resource Locator ) est l'adresse standard utilisée pour chercher et récupérer des documents en World Wide Web.

##### **WWW englobe trois composantes:**

**HTML** : un format de description de documents hypertexte, multimédia, multiprotocole;

**HTTP** : un protocole permettant la récupération de documents HTML localisés sur des serveurs;

**WWW** : un réseau mondial de serveurs délivrant des documents en format HTML.

Il apporte un nombre appréciable d'améliorations sur plusieurs plans:

- **ergonomie** : on parcourt Internet en cliquant sur des mots situés dans des textes;

- **autonomie**: on ne dépend plus d'un serveur et chacun peut, avec son éditeur de textes favori, rédiger un texte HTML correspondant à sa vision du monde ou d'un problème, le lire, en exploiter les liens, l'envoyer à ses correspondants par messagerie ou FTP, ou enfin le stocker sur un serveur WWW;

- **fonctionnalités** : un lien HTML peut pointer vers un nombre d'objets variés (fichier, espace FTP, news-group, index wais, serveur Gopher, serveur WWW, etc...)



### 3.3.- Description et structure du Serveur Handy:

#### 3.3.1.- Objectif:

Ce serveur est destiné aux étudiants handicapés, et aux enseignants ayant des étudiants handicapés dans leurs cours. Ils y trouveront les informations relatives aux services, aux équipements, aux marches à suivre aux dispositions légales concernant, le tiers temps, etc... Noter cependant que la mise en place de ces rubriques se fera au fur et à mesure des possibilités.

Ce serveur permet également d'accéder à divers autres services dans le monde Internet en rapport avec des besoins des handicapés ou des personnes intéressées.

#### 3.3.2.-Description:

Le serveur de la Mission propose les rubriques suivantes:

##### a) Services offerts:

- **Cours et documents** : texte intégral en ligne ((accès réduit à la Mission Handicap et aux personnes autorisées).

Cours de deug Charbonnier

TP d'informatique DEUG A Chimie des matériaux

Maths DEUG intègre. 1er semestre

Moisissures - Mycotoxines et Aliments : du risque à la prévention.

Almanach anti-superstitieux de Condorcet

Cours de C par G.-H. Lamot

- **Soutien pédagogique** : prise de notes

Certains types de handicaps comme les déficients auditifs profonds exigent qu'une personne valide prenne les notes de cours. Cette personne est à trouver parmi les étudiants de la même formation, une rémunération est prévue (environ 30 fr. de l'heure).

Des conditions spécifiques sont requises pour effectuer les prises de notes: avoir la nationalité française ou bien avoir une carte de séjour qui autorise un travail rémunéré. Les preneurs de notes seront choisis en accord entre l'étudiant handicapé, les enseignants et la mission handicap. Il est fortement recommandé que chaque enseignant valide les notes prises. Lorsque des photocopies sont nécessaires, une carte de photocopie est fournie gratuitement à l'étudiant handicapé, à la DOUA les photocopies peuvent être faites directement à la mission handicap (d'ambulatorio 1er cycle, entre amphi 2 et amphi 3).

### **- Soutien pour le passage des examens:**

Circulaire du ministère N° 4 du 22 mars 1994

Détaille les conditions de passage des examens. Il est possible maintenant d'utiliser des micro-ordinateurs avec l'accord du jury. Il doit être tenu compte du tiers temps quand plusieurs épreuves se suivent. Autre point important, la compétence des secrétaires est définie. Une certaine souplesse est accordée, qui permettra à notre Université d'adapter ce qui se passe habituellement aux conditions requises.

La mission handicap, selon son budget, peut programmer des vacations pour des heures de surveillance et/ou de secrétariat. Le responsable du jury, devra prévenir Mr DECORET en temps voulu, tant pour permettre la mise sous une forme adaptée des sujets (transcription braille) que pour résoudre les modalités pratiques de rémunération.

### **- Accès Bibliothèque pour les non voyants:**

Le local informatisé pour les étudiants handicapés est situé à droite en entrant, auez de chaussé de la Bibliothèque Universitaire (campus de la Doua). L'accès de cette salle est limitée aux personnes autorisées par la mission handicap. Pour l'inscription contacter le responsable.

L'équipement comprend une machine à lire READING EDGE, un micro-ordinateur 386 adapté avec une synthèse vocale permettant une lecture en français, anglais, allemand ou espagnol des textes qui apparaissent sur l'écran du micro-ordinateur (marque: APOLLO), un pilote d'écran pour la synthèse vocale (SCREEN READER), une imprimante permettant de sortir des textes en braille (marque: INDEX), une imprimante ordinaire, une carte réseau permettant l'accès des étudiants à Internet, un logiciel de lecture de documents cryptés protégés par droits d'auteurs (livres de BRAILLESOFT, logiciel HYPERDIFF), des disquettes d'ouvrages dont 17 Que Sais-je et quelques livres d'informatique.

Un logiciel de grossissement de caractères à l'usage des malvoyants (ZOOMTEXT)

Une loupe électronique

### **- Accès au réseau informatique Internet**

- ® Introduction aux outils du réseau
- ® Sommaire
- ® Connexion au réseau
- ® Le mail ou le courrier électronique
- ® La commande pour parler inter activement à l'interlocuteur
- ® La command FTP
- ® Le vi, un éditeur puissant de unix

## **- Information sur les matériels d'adaptation et des fournisseurs**

Matériels d'adaptation vendus par l'Association Valentin Haüy (à l'usage des aveugles et mal-voyants)

### **- ASSOCIATION VALENTIN HAÜY POUR LE BIEN DES AVEUGLES**

5 rue Duroc 75343 PARIS cedex 07

Télé: (+33) 1 44 49 27 37

Fax: (+33) 1 44 49 27 20

### **Catalogue Menu:**

- Brailscope
- Memona 4956
- Ecrans braille:
  - écran braille 4917
  - écran mini braille
- Imprimantes:
  - imprimante index
  - imprimante index basic 8 points
  - imprimante braille portal-thiel
- Logiciels CD ROM
- Scanner:
  - Système de lecture Arkenstone Reader 2
- Divers
  - Machine à lire Xerox

### **- Plan d'accès de la Mission Handicap:**

Université; Claude Bernard - Lyon I \_UCBL

Plan du campus

**a)** Consultable par morceaux (pour les petits écrans)

**b)** Consultable en une pièce (pour les grands écrans)

### **- Annuaire téléphonique du campus**

## **b) Autres accès**

### **Accès spécialisés**

#### **- Association ABU:**

L'Association de Bibliophiles Universeles: : C'est une association loi de 1901 fondée en Avril de 1993. Elle a pour objectif le développement et la promotion des supports numériques permettant la libre manipulation de l'information, l'application de ces techniques à la diffusion des travaux de recherche des membres et d'informations du domaine public.

Cette association se propose de donner un maximum de textes du domaine public en langue française, qu'il s'agisse de textes anciennes ou de création. ABU diffuse aussi une copie des textes anglais du Project Gutenberg et héberge les pages de quelques organismes encore plus pauvres qu'elle.

#### **- Poèmes**

#### **- Université de St. Johns:**

Serveur gopher de la St John's university Academic Computing (Etats Unis). Il a des informations relatives aux handicapés; groupes et organismes, sources sur l'éducation et technologie, entre autres.

#### **- Catalogues des livres complets en ligne accessibles par Internet**

#### **- Widnet**

Word Institute of Disability : Informations intéressantes pour les handicapés

#### **- Serveur ftp Handicap**

#### **- Enregistrements sonores (cassettes) pour les non voyants**

#### **- ADA : American With Disabilities Act**

Serveur avec des informations relatives aux handicapés, exemple: Americans with disabilities Act of 1990, Americans with disabilities Act Accessibility Guideline, Americans with disabilities Act documents Cents; Social security Administration- Disability Benefits, etc.

#### **- Trace Center ( Sourds - muets)**

#### **- Le programme d'aide de DO-IT.**

Il offre une liste de mailing-list, des circulaires des news groups et des sites gophers qui contiennent des informations intéressantes pour les personnes handicapées.

- **Le serveur Web de DO-IT**
- **Royal National Institute for the Blind** : RNIB Association anglaise pour aveugle.
- **Autres adresses spécialisées.**

### *Journaux électroniques français*

- Le Monde Diplomatique
- Le Cahier Multimédia de Libération
- Radio Prague en Français
- Cybersphère : le magazine électronique de la Cyberculture.

### *Surdité*

- Deafness Ressources
- Deaf World Wet

### *Les News*

- Newsgroups . bit-listserv.1 hcap et misc handicap, avec des informations intéressantes pour les handicapés.

## **c) Autres accès par le monde**

### **Les outils pour rechercher des informations sur les serveurs Web**

- Web Crawler Database
- Yahoo Search

### **Autres**

- O'Reilly & Ass Global Network Navigator
- Tous les Gophers
- Librairie du Congrès US
- Journaux Etrangers

#### **d) Serveur Wais Expérimental:**

- **Base générale de la Mission handicap**, cette page permet d'accéder par mots-clés à l'information contenue dans le serveur Web

- Mots clés
- Type de recherche : recherche sur le site WWW Hardy résultat complet.
- Recherche sur le site WWW Hard (résultat avec mot clés en gras)
- Option Wais qui ne tourne pas

#### **e) Utilitaires:**

- Description française de Wais
- Pour naviguer à partir d'un terminal ASCII avec l'utilisation du logiciel Lynx.

## **IV. - UTILISATEURS DU SERVEUR HANDYNET:**

### **4.1- Usagers:**

Le serveur Handy a comme objectif satisfaire les besoins d'information des élèves handicapés de l'Université Claude Bernard Lyon 1. L'univers des utilisateurs à niveau réel est de 40 étudiants qui réalisent leur formation supérieure dans le domaine scientifique et technique en accord avec les diplômes donnés par l'université.

Dans ce cadre les documents constituant la base de données, sont des documents sur format électroniques de cours et TD faites pour les enseignements de l'université. La couverture thématique des documents à inclure dans la base se trouvent principalement dans les domaines des sciences, sciences appliquées, sciences de la vie, sciences de la terre, sciences pour l'ingénierie; comme par exemple: biologie, écologie, physique, chimie, biochimie, santé, pharmacie, génie mécanique, informatique, électronique, génie civil, mécanique, mécanique énergétique, entre autres.

La Mission a mis aussi à disposition des utilisateurs, des documents et d'information de type complémentaire, soit de formation générale (poésie, livres en hypertexte, journaux en français, etc.) ou soit des informations intéressantes pour les handicapés, comme par exemple : information relative au Royal National Institute for the Blind, Catalogue de livres complets en ligne accessibles par Internet, etc.

## 4.2.- Les Formations à l'Université Claude Bernard

### La formation initiale à l'UCBL

L'Université Claude Bernard - Lyon 1 est une université scientifique et médicale qui propose un large éventail de formations préparant à l'entrée dans la vie professionnelle, aussi bien dans les secteurs industriels et tertiaires que dans l'enseignement ou les professions libérales. Parmi les études qu'elle propose, certaines permettent de s'orienter vers une formation par la recherche, ouvrant à des emplois dans la recherche, publique ou privée, mais aussi dans l'industrie ou dans l'enseignement supérieur. La même variété se retrouve dans les disciplines enseignées:

Le secteur **SCIENCES** propose des formations dans les matières fondamentales (mathématiques, physique, chimie), les sciences de la vie (biochimie, biologie, écologie, physiologie) et de la terre, les sciences pour l'ingénieur (technologie industrielle, informatique, électronique, génie civil, génie biologique et médical; l'université est notamment habilitée depuis 1992 à délivrer un titre d'ingénieur en multimatériaux, mécanique, productique et mathématiques appliquées) dans le cadre de l'ISTIL.

Le secteur **SANTE** recouvre les études de médecine (générale et spécialisée), odontologie, pharmacie (officine, industrie, pharmacie hospitalière ou biologie médicale), les formations en biologie humaine (maîtrise de sciences biologiques et médicales), mais aussi les techniques de réadaptation dans le secteur paramédical (kinésithérapie, ergothérapie, audioprothèse, orophonie, orthoptie, psychomotricité).

Les formations assurées par les Instituts Universitaires de Technologie ouvrent aussi bien sur l'industrie de la biologie appliquée, au génie mécanique et à la maintenance industrielle, en passant par la chimie, le génie électrique et le génie civil, que sur le secteur tertiaire (gestion, commerce), secteur lui-même couvert par d'autres formations, présentes en Sciences ou en santé.

### Les études doctorales à l'UCBL

Le 3e cycle de l'enseignement supérieure comprend:

- Des études courtes ( 1 an) à dominante professionnelle débouchant sur le

#### - Diplôme d'Etudes Supérieures Spécialisées ( D.E.S.S).

**Diplômes de D.E.S.S.** : Electronique, électrotechnique et automatique; génie biomédical; images de synthèses, principes et applications; informatique documentaire; informatique, image et productique; ingénierie mathématique; microbiologie; pharmacie industrielle; gestion de la qualité appliqué à l'industrie chimique et à l'industrie du médicament

- Des études longues (3 à 4 ans) permettant la préparation d'un doctorat grade le plus élevé décerné par l'université française - après l'obtention, la première année, d'un Diplôme d'Etudes Approfondies (D.E.A.).

*Diplômes d'études approfondies D.E.A.:* Analyse numérique modèle mathématiques calcul scientifique; mathématiques; astrophysique; physique statistique et milieux dilués; matériaux macromoléculaires et composites; sciences de matériaux et des surfaces; analyse Physico-chimiques chimie analytique; chimie organique fine; physique théorique; sciences de la terre et de l'atmosphère; paléontologie dynamique sédimentaire, chronologie; informatique fondamentale; mécanique; ingénierie informatique; génie civil, sols, matériaux, structures physiques d'université Bâtiment; automatique industrielle; génie électrique; sciences et technologie industrielles de la chimie; thermique et énergétique; dispositifs de l'électronique intégrée; acoustique; cinétique et catalyse, du matériau catalytique au procédé industrielle; adaptation et survie en environnements extrêmes; national de toxicologie; pharmacocinétique fondamentale et expérimentale et effets du médicament chez l'homme; biologie et pharmacologie cutanées; génie biologique et médical; écologie microbienne; neurosciences; neuropsychologie; biologie de l'exercice musculaire; métabolismes, régulations endocriniennes, nutrition, développement; biomécanique de l'appareil locomoteur et biomatériaux de remplacement; parasitologie; biochimie; pharmacologie cardio-vasculaire; analyse et modélisation des systèmes biologiques; différenciation génétique et immunologie; didactique des disciplines scientifiques; sport, performance et environnement social; psychologie et psychopathologie cliniques; méthode d'analyse des systèmes de santé;

## **Formation Continue et alternance Lyon 1**

Le Service Commun de la formation continue a pour mission générale de favoriser le développement des activités de formation continue et d'éducation permanente que toutes les composantes de l'Université sont en mesure de proposer à des partenaires des secteurs industriels ou des services ainsi qu'aux demandeurs individuels.

### *- Formation qualifiante non diplômante:*

Des stages mis en place des équipes de recherche ou d'enseignement pour répondre à des besoins qui ont été décelés ou en réponse à des demandes spécifiques d'entreprises ou organismes. Ils n'ont pas de caractère diplômant et sont financés par les employeurs ou les individuels, notamment lorsqu'ils concernent les professions libérales. Ils permettent aux professionnels d'actualiser leurs connaissances liées aux derniers résultats de la recherche avec des possibilités d'accueil dans les laboratoires.

### *- Formation diplômante:*

Des stages longs, le plus souvent diplômants (diplômes nationaux ou diplômes d'Université fréquemment homologués par le Ministère du travail, d'emploi et de la formation professionnelle), spécifiques au public de la formation continue ou non spécifiques avec intégration dans les cursus de la formation initiales, à temps pleins ou à temps partiel, avec des possibilités individualisés.



## **La recherche à l'Université Claude Bernard Lyon 1**

- L'Université couvre de très nombreux champs disciplinaires du fait de sa taille et de sa bipolarité Sciences-Santé. Les recherches qui s'y poursuivent ont une caractéristique essentielle: elles sont reconnues dans leur quasi-totalité par les grands organismes nationaux. Ainsi la plupart des laboratoires sont-ils soutenus par le CNRS, l'INSERM, ou la Mission scientifique et Technique du Ministère de l'Equipelement Supérieur et de la Recherche.

- La fonction spécifique d'une recherche universitaire étant la part importante consacrée à la formation des jeunes chercheurs - qui iront ensuite enrichir les laboratoires publics et privés - l'Université Claude Bernard remplit largement son rôle puisqu'elle pilote directement près d'une trentaine des formations doctorales et qu'elle est associée à une cinquantaine d'autres. La recherche à l'université se distingue par quatre aspects originaux:

a) Sa pluridisciplinarité, exemplaire dans le domaines comme l'environnement, la biomécanique et les biomatériaux, l'imagerie médicale, les recherches sur le sport, l'économie de la santé, les sciences actuarielles, les groupements de recherche comme l'Ecole Interdisciplinaire sur le Médicament (EDIM).

b) Sa percée significative dans les secteurs avancées: agrégats d'atomes, lasers, physique des particules, matériaux, génie des procédés - en lien avec la nouvelle école CPE - calcul massivement parallèle - en lien avec l'INSA.

c) Son soutien à des regroupements géographiques de laboratoires, afin de constituer des entités plus compétitives sur le plan international, en génétique moléculaire, en biochimie, chimie, mathématiques.

d) Son association active à des opérations nationales et internationales importantes: création de l'Institut des Cognisciences de CNRS, réalisation de l'Institut Fédératif des Neurosciences par l'INSERM, projet européen VIRGO.

## **V.-PROCEDURE DE COLLECTE D'INFORMATION ET TRAITEMENT DES DOCUMENTS ELECTRONIQUES:**

### **5.1.-Gestion électronique de documents:**

#### **a) Qu'est-ce qu'un document <sup>1</sup> :**

- L'étymologie fait référence au verbe latin "docere" qui signifie instruire.
- L'étude de la langue française ancienne remémore qu'un document est "ce qui sert à instruire".

---

<sup>1</sup> LASSOURY, Jacques. La documentique : gestion électronique de documents et gestion documentaire. Paris : DUNOD, 1994. 154p.

- Les dictionnaires contemporains s'accordent sur une définition quasi-consensuelle: "renseignement écrit, servant de preuve ou d'information".
- L'ISO de son côté déclare: "un document est l'ensemble d'un support d'information et de données enregistrées sur celui-ci sous une forme générale permanente et lisible par l'homme ou par une machine.

## **b) Les usages de documents:**

De nombreux travaux ont en particulier ceux menés par l'AFNOR ont cherché à définir une classification des documents en fonction de leur usage. Un quasi-consensus est apparu sur une classification selon 7 rubriques : Documents personnels et individuels - Documents administratifs et de gestion - Documents de l'état et de la puissance publique - Documents techniques et scientifiques - Documents commerciaux et bancaires - Documents financiers et comptables - Documents religieux.

D'accord à la multiplicité des définitions proposées, une approche pragmatique du concept même de document consiste à définir le document en fonction de son usage, de son support et de son contenu. En ce qui concerne l'activité de la Mission on parle alors, de la gestion électronique de documents électroniques (support) de type scientifique (usage). Au sujet du contenu, la plupart des documents que l'on manipule quotidiennement sont de type texte (sous formats word, wp, rtf, ascii, etc.) et ils sont souvent accompagnés des formules mathématiques, des formules chimiques et de tableaux.

La Mission dispose d'une collection de documents électroniques de type scientifique et technique comme des cours, des TD, des documents de soutien pédagogique entre autres et il ont les caractéristiques déjà définies selon son support, contenu et usage. Par rapport à la gestion des documents au sein de la Mission, on a qu'elle mise en place une série des procédures grâce à l'installation d'outils informatiques (words pour windows), qui permettent gérer les documents mis sous forme des fichiers HTML.

En relation au traitement et l'utilisation du document structuré, la Mission a fixé des règles de présentation dans le traitement de texte, qui permettent de différencier au sein du document les données de la structure (les relations entre les données) et du format (police de caractères, marges, etc.) à travers d'un balisage logique d'accord aux paramètres et critères de HTML 2.0.

## **c) La collecte d'information:**

### **Sélection :**

Le processus de collecte d'information au sein de la Mission est fait par un spécialiste qui prend contact personnel avec les professeurs universitaires, son objectif est de déterminer avec eux les documents susceptibles d'être mis sous le serveur. Les textes reçus peuvent être écrits avec différents outils de traitement de texte (word, wp, etc.) sur PC, Mac ou sur une machine unix.

La Mission prend contact aussi avec différentes entités, avec pour objet de connaître la production intellectuelle des enseignants:

a) Associations des Etudiants (eux signalent les documents qui sont mis à disposition par les professeurs.

b) IREM (Institut de Recherche en Mathématiques)

Il faut détacher que la Mission a aucun contact avec les services de reprographie de l'université et ne peut donc pas connaître à niveau réel la production des textes scientifiques et techniques. Dans ce cadre on a fait une enquête dirigée aux trois services plus importants de reprographie, afin de définir d'une manière systématique "le processus de collecte d'information" et de délimiter les tâches au sein de la Mission pour améliorer la chaîne documentaire. Malheureusement, l'enquête n'a été pas répondue de façon formelle, mais l'information obtenue a montré que dans le sein de l'université n'existe pas encore une politique de diffusion et de dissémination massive de documents au niveau de soutien pédagogique, tant à niveau de diffusion en systèmes d'information traditionnels, comme à niveau de dissémination en ligne (réseau).

Un point important dans la réalisation de ce stage, a été étudier tout le processus de collecte d'information et chaîne documentaire. Après le contact avec le spécialiste qui réalise la sélection des documents, l'étude du serveur, l'analyse du balisage HTML et les problématiques détectées, on a arrivé à la conclusion qu'il y a deux problèmes fondamentaux:

a) Collecte d'information: Ce processus n'est systématique et n'existe pas interaction entre les services de reprographie, la bibliothèque, les enseignants et la Mission Handicap; problème qui a eu une incidence dans la procédure de sélection des documents et la postérieure problématique de pertinence / complétude de l'information mise dans le serveur.

b) Réalisation de Balisage HTML: La procédure inclut des tâches relatives à conversion des fichiers formats texte (word, Wp, etc.) à un format standard (RTF) et postérieur balisage à langage HTML. On trouve des problèmes par rapport à la reconnaissance des caractères spéciaux, reconnaissance des formules mathématiques et chimiques. Dans ce cadre, la problématique fondamentale à résoudre est de type technique; c'est-à-dire: il faut trouver dans la vieille technologie, la solution à problèmes relatifs à la conversion des fichiers (trouver des filtres et des convertisseurs appropriés pour le balisage HTML), arriver à la reconnaissance des formules et améliorer la procédure de balisage, parce que la tâche prend beaucoup de temps et il n'y a pas de personnel suffisant pour réaliser l'activité.

#### **d) La Chaîne Documentaire:**

##### ***Procédures***

- **Acquisition** :contact avec les professeurs, contacts avec les élèves, contact avec l'IREM, sélection du documents, obtention du texte en version disquette.

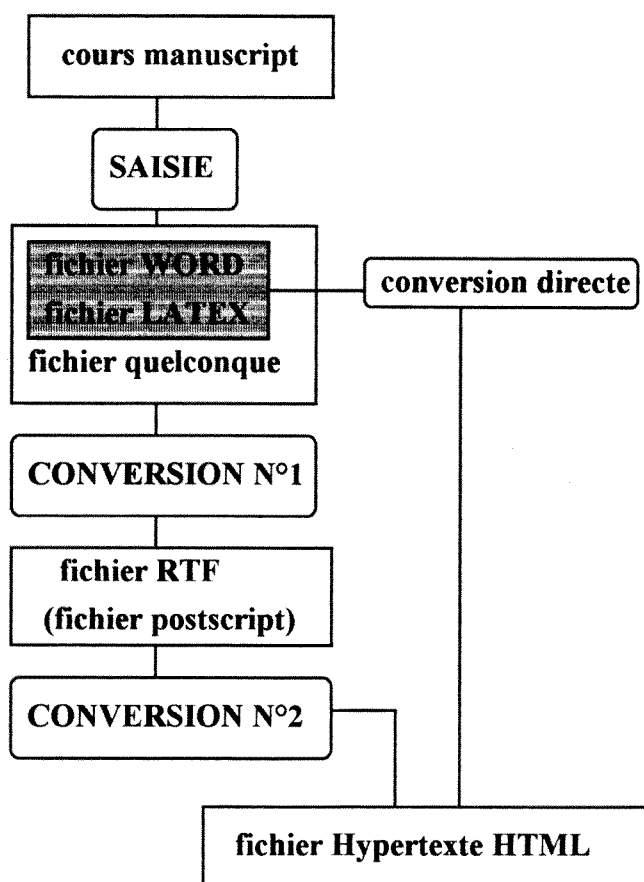
- **Conversion des fichiers:** Wp, word, ascii ----- à fichier RTF ----- (tâches de retouches, validation, annotation, modifications) ----- à fichiers sous norme HTML 2.0

- **Archivage ou stockage** : réalisation de balisage HTML ( marquage, style, liens)

- **Indexation** (indexation des fichier HTML)
- **Circulation** (disposition du document en hypertexte dans le serveur)
- **Consultation** : affichage, navigation au sein du document, recherche.

## MISSION HANDICAP

Chaîne de traitement pour l'installation de documents sur le serveur



### Saisie du cours:

Cette opération peut être effectuée soit par l'enseignant, soit par une personne spécialisée de la Mission Handicap.

### Recommandations aux auteurs :

Les applications spécialisées qui dessins de formules chimiques, dessins d'

## **e) Récupération et structuration HTML de documents:**

Le travail de la Mission Handicap consiste à fournir des documents SGML ( norme incluant les notations scientifiques) et les mettre à disposition des déficients visuels.

### **\* Du papier à HTML en passant par l'ASCII**

La structuration automatique de fichiers ASCII nécessite une analyse automatique de structure (reconnaissance de paragraphe, par exemple) à l'aide de logiciel comme Markit. Les documents ASCII peuvent être ceux obtenus par reconnaissance optique et reconnaissance de caractères. Pour cela la Mission a une machine à lire très performante et récente (XEROX)

La récupération par ces systèmes exige également une phase de correction qu'il est difficile d'automatiser et consommatrice de temps. Cette même phase implique une reprise manuelle des expressions mathématiques avec le langage HTML. Les documents sont stockés sur une base de données et sont utiles aux étudiants handicapés de l'UCBL, ainsi qu'aux enseignants.

### **\* Tex (ou latex) à HTML**

Tex est un langage d'écriture de textes scientifiques. la Mission a créé un ensemble de logiciels permettant d'obtenir à partir de textes écrits en TEX des impressions braille mathématiques. Ce travail est une préparation pour placer des marques SGML à partir d'un texte écrit en Tex.

### **\*wp, word... (les traitements de texte) à HTML**

Certains de ces traitements de texte présentent des possibilités pour placer des marques HTML (SGML étendu à la notion d'hypertexte). La Mission a déjà utilisé ces facilités dans la mise en place du serveur WWW (<http://handy.univ-lyon1.fr>). Il s'agirait d'utiliser ces facilités pour la pose automatique de marques en particulier pour les signes mathématiques et scientifiques susceptibles de pouvoir suivre un traitement adapté pour le braille ou les sorties vocales.

## VI. - BALISAGE HTML

### 6.1.- Introduction à SGML<sup>2</sup> :

La norme internationale SGML (Standard Generic Markup Language) traite du balisage structurel des documents sous leur forme électronique. Elle fut adoptée par l'ISO en Octobre 1986 et son importance est bien reflétée dans le fait que c'est une des rares normes ISO traduites en français. Elle est vite devenue très populaire, même en Europe, notamment par son acceptation dans le monde des éditeurs, des grandes entreprises, des institutions gouvernementales, etc., plus récemment, par le biais de HTML, le langage hypertexte de W3.

Les points d'un balisage SGML sont:

- une amélioration de la qualité des sources des documents;
- une rationalisation dans le traitement du document, due en grande partie à un cycle de travail plus rapide;
- une réduction des coûts des publications;
- la possibilité de réutilisation de l'information, d'où une plus-value (imprimés, hypertextes, bases de données).

Une représentation de documents basés sur la description de leur structure logique est essentielle pour les traiter électroniquement. Pour garantir que les documents soient réellement échangeables on a dû mettre d'accord sur un langage commun. C'est qui fut fait lors' en Octobre 1986, ISO adopta officiellement SGML; le langage standard de balisage généralisé, comme norme internationale ISO 8879.

#### 6.1.1.- Principes de balisage SGML:

SGML est une méthode normalisée permettant de représenter l'information contenue dans un document indépendamment des systèmes utilisés pour la saisie ou son traitement, et indépendamment de la forme physique finale.

SGML repose sur le principe de balisage logique des documents et met en oeuvre ce principe sous la forme d'un balisage logique généralisé. SGML n'est pas seulement un langage de balisage, mais un outil pour construire plusieurs langages de balisage, autrement dit un métalangage.

---

<sup>2</sup> GOSSENS, Michel. Introduction pratique à SGML. EN: Cahier Gutenberg. Rennes, France : IRISA.

Janvier 1995 (19): pp. 27-58.

## 6.2.- Utilisation de HTML à la Mission Handicap:

### \* Généralités

**Html (Hypertext Markup Langage)**<sup>3</sup> a été conçu comme un classe de document SGML, norme ISO 8879 et est assimilable à un langage, basé sur un ensemble de balises permettant de décrire des documents d'une manière structurée. Les différents structures présentes dans un documents HTML sont encadrées par un paire de balises : un balise de début **<Non d'élément>** et un balise de fin **</Non d'élément>**. Bien que la présence de certaines balises de fin ne soit pas obligatoire, il est souhaitable qu'à chaque balise de début corresponde une balise de fin.

Comme dans tout documents SGML, les nom de balises peut être indifféremment écrit en minuscules, en majuscules avec un combinaison des deux. de plus, plusieurs espaces successifs interrompus ou non par des changements de ligne sont pris en compte comme un seul espace et les caractères accentués ne doivent pas être utilisés directement.

Tous les documents HTML doivent posséder une structure minimale commune divisée en deux parties distinctes:

- un en - tête, où sont données des renseignements d'identification;
- un corps, où se trouver la partie visible (texte du document)

Ces éléments de structure du document **HTML** sont marqués par les balises **<HEAD>** et **</HEAD>**, qui encadrent l'en-tête, et **<BODY>** et **</BODY>**, qui encadrent le corps. L'ensemble doit être encadré par les balises **<HTML>** et **</HTML>**. L'en tête doit contenir au moins un élément, **TITLE**, faisant office de titre **<externe>**<sup>4</sup> par le document.

### \* Structures usuelles

Les structures autorisées en **HTML** sont de plusieurs ordres. Certains servent à diviser les texte en parties distinctes, d'autres servent à indiquer des préférences de mise en forme de texte en enfin d'autres permettent d'utiliser des fonctions spéciales.

---

<sup>3</sup> ROLLAND, Christian. Présentation de HTML. EN: Czahier Gutenberg. Rennes, France: IRISA.

Janvier 1995 (19): pp. 67-84.

<sup>4</sup> Le titre externe n'est pas le titre visible lorsque l'on consulte le document. Il s'agit d'une manière de faire référence à ce document, un peu comme le nom d'un fichier, même si ce titre externe peut être plus explicite. Un document peut donc posséder deux titres à la fois : un titre externe ne faisant pas partie du texte du document et un titre interne inclus dans le texte du document.

## **\*\* Division structurelle du document:**

Dans HTML, le document n'est pas explicitement divisés en parties, chapitres, sections, etc.; comme c'est le cas d'autres langages. la division se fait par une hiérarchie d'éléments dont chacun est libre de donner la signification qu'il désire (titre de document, section, sous-section, titre de paragraphe, etc.) En effet, le titre de ces éléments de division est précisé selon son niveau dans une hiérarchie : les balises son nommées *Hn* où *n* est le niveau dans la hiérarchie : *H1, H2,....., H6*.

## **\*\* Listes:**

Cinq types de listes sont disponibles dans HTML. Ces listes sont délimités par des balises qui indiquent le type de liste représenté. Chaque élément constituant la liste balisé <LI> et </LI>;

### *Types de listes*

- **Liste ordonnée:** Elle est représenté par l'élément **OL**. Il s'agit d'une énumération d'éléments. Le compteur est affiché au début d'un nouvelle ligne, suivi de l'objet de liste correspondant.

- **Liste non ordonnée:** Une liste non ordonnée est représentée par l'élément **UL**. Elle est composée d'une série d'éléments non ordonnés (aucun numéro d'ordre ne leur est associé).

- **Liste descriptive:** Une liste descriptive est représentée par l'élément **DL**. Contrairement aux listes précédents, les objets d'une liste sont composés de deux parties: la partie entrée et la partie description. L'entrée est représenté par l'élément **DT** et la partie description par l'élément **DD**.

L'attribut *COMPACT* peut-être associé à l'élément glossaire en vue d'une disposition de la liste, si possible plus <<compact>> (par exemple, placement de l'entrée en face de la première ligne de la description au lieu de la placer sur une ligne seule).

- **Liste de type répertoire:** Cette liste est représenté par l'élément **DIR**. Elle doit être composée d'éléments contenant chacun moins de 20 caractères. Selon les logiciels utilisés, les listes de ce type pourront être ou non disposées en colonnes.

- **Liste type menu:** Une liste de type menu est représenté par l'élément **MENU**. Elle est composée d'éléments dont la longueur ne dépasse pas une ligne. Chaque objet de la liste est généralement placé au début d'une nouvelle ligne lors de la mise en forme du document pour la visualisation.



### **6.3.- Indexation et recherche de documents:**

Les documents électroniques en hypertexte enregistrés dans la base de données sont repérés à l'aide de descripteurs appelés "index" ou "attributs". ceux-ci peuvent être contenus, soit dans le documents, soit constituer une information externe au document. L'indexation se fait par un opérateur de façon manuelle.

### **6.4.- Recherche de documents:**

Les modèles de recherche sont:

- Recherche directe: Se fait à partir des descripteurs externes qui sont le nom ou la référence du document;
- Recherche par mots-clés : la recherche se fait à l'aide d'index inclus dans le document de référence qui comprend des informations utiles à la recherche documentaire.
- Recherche en texte intégral: Ce type de recherche est une extension des méthodes des recherche par mots clés. dans ce cas le document de référence est constitué par l'ensemble du document; ses méthodes permettent de retrouver un document à partir d'un élément quelconque de son contenu.
- La navigation hypertexte: on a la possibilité de naviguer au sein d'un ensemble de documents en déboisant sur des mots.

## **VII.- ETUDE DES BESOINS:**

### **7.1.- INTRODUCTION:**

Le serveur de la Mission Handicap a des documents électroniques de type cours, des Tp, des documents de type soutien pédagogique. La problématique principale de la mission concerne le fonds de la base de données disponible ; y a-t-il un décalage entre la production des documents format papier et les documents électroniques des cours donnés, TD , etc. Les sources existantes dans le serveur sont insuffisantes pour satisfaire les besoins d'informations spécifiques de tous les usagers du serveur de la Mission dans ce cadre, il y a la possibilité de faire un étude permettant de connaître les besoins et renseignements des utilisateurs. Il faut étudier aussi, les documents disponibles dans le serveur et sa pertinence avec l'objet de complémentaire ce système d'information en ligne avec un service de référence en ligne.

Ce dernier est orienté pour fournir l'information sur les documents pertinentes, qui intéresse les utilisateurs. Ces sources seraient sélectionnés à partir des différentes bases de données accessibles par les serveurs d' INTERNET.

La tâche de collecte d'information doit être orienté pour satisfaire les besoins spécifiques liés à les cours donnés à Université et à tous les activités de recherche.

Dans ce qui concerne au traitement de documents électroniques, il faut étudier tout le processus afin de déterminer les tâches à améliorer.

## **7.2.- OBJECTIFS**

### **7.2.1.- Objectif Principal :**

Améliorer le niveau d'utilisation réel du serveur de la Mission et mettre à disposition des usager l'information pertinente. Ce qui permettra de les aider à résoudre de façon efficace la problématique de recherche d'information dans les différents cadres ou activités académiques de l'Université Claude Bernard Lyon 1.

### **7.2.2.- Objectifs spécifiques:**

- a) Compléter l'information scientifique et technique disponible dans le serveur, avec la création d'un guide en ligne de sources et/ou documents électroniques en français, disponibles sur autres serveurs tant qu'en France qu'à l'étrangère.
- b) Diffuser les documents électroniques existants dans le cadre scientifique et technologique, qui sont accessibles au travers le serveur de la Mission.
- c) Optimiser l'utilisation des documents existants dans la mission, grâce à la création d'un catalogue des documents électroniques, qui permet de faire de recherches par critères ou champs spécifiques où l'utilisateur pourra obtenir une liste de documents pertinents. Eventuellement l'utilisateur pourra obtenir les références pertinentes et faire un lien HTML pour accéder aux documents hypertexte.

### **TACHES : (Pour réaliser objectifs 1 et 2)**

- a) Faire une étude relative à l'utilisation réelle du serveur de la mission au travers d'une enquête dirigée vers les utilisateurs, avec l'objectif de savoir si les documents électroniques disponibles satisfassent ses besoins d'information.

Grâce à cette enquête on pourra connaître aussi les faiblesses du serveur, par rapport à la quantité de documents disponibles (problème de complétude) et la qualité des documents (problème de pertinence). En même temps, on peut s'apercevoir des problèmes relatifs aux critères de recherche d'information (pertinence de documents en recherches de type hypertexte), serveur amiable, commandes de recherches faciles, ....

- b) Etudier la chaîne de collecte d'information de documents réalisés à la Mission.
- c) Etude de documents existants. Lister les documents dans le serveur et collecter tous les documents électroniques faits par les enseignants de l'Université.
- d) Réviser les documents électroniques de types scientifiques et technologiques disponibles dans le serveur Handinet.

- e) Repérer les entités connectées à INTERNET ayant des documents dans les matières étudiées par les élèves handicapés de l'Université, et puis, étudier les documents électroniques en français disponibles dans les différents serveurs.
- f) Exploitation du annuaire "Golden Directory of INTERNET" (dernier édition), avec l'objectif de connaître par sujet les sources d'intérêt pour les utilisateurs de la Mission et après inclure les sources obtenus dans le guide de référence en ligne.
- g) Création d'un guide en ligne sur les sources disponibles et de documents d'intérêt pour les usagers de Handinet, et faire sa réalisation en accord avec les paramètres de stockage SGML.
- h) Saisie de l'information en format word et passage à fichier sous SGML pour balisage.
- i) Liaison avec les autres serveurs.

### **TACHES A REALISER POUR OBTENIR OBJECTIF 3:**

- a) Etude de la collection de documents électroniques existants dans le serveur Handinet, avec l'objectif de définir la structure de données factuelles à décrire (liste de champs).
  - b) Création du catalogue de documents : quel type de catalogue: auteur, matière, auteur. et sur quel support.
  - c) Réalisation de registre bibliographique (structure de notice, format et contenu)
  - d) Saisie de l'information ou de notices.
  - e) Conversion de ce catalogue écrit en format word à une structure SGML et sa mise à disposition sur le serveur Handinet.
  - f) Faire une liaison entre la notice bibliographique et le document hypertexte.
  - g) Liaison indirecte : donner l'adresse du documents hypertexte dans la structure de la notice (création de un champ).
- Ou , liaison directe en langage HTML au document hypertexte.

### **7.3.- Analyse des applications à réaliser:**

Le responsable de la Mission Handicap a étudié , toutes les propositions faites par la stagiaire, avec l'objectif de déterminer quelles sont les priorités ou activités à privilégier.

Le résultat de cette analyse est le suivant:

- a) Enquête: Il y actuellement peu d'utilisateurs, donc une enquête proprement dite est prématurée. Cependant, il est important de mettre au point des recommandations sur le type d' enquête.

L' exploitation du logiciel "espion" (daemon http) est indiquée , si ce ne serait pas suffisant dans l'avenir, mais comment analyser.

b) Il est important de connaître la collecte et donc d'étudier la chaîne de collecte d'information (récupération de disquette, documents papiers). La recommandation de C.E.V.U. dit que a chaque polycopie donné a la reprographie (l'un des trois centres) passe par le dépôt d'une disquette. Faire une enquête sommaire pour savoir si la demande est pertinente.

c) Il faut lister les documents sur le serveur et les documents qu'on devrait avoir.

d) Etudier les documents électroniques de type scientifiques et technologiques en mettant en lumière les caractéristiques de documents.

Tâches pour réaliser l'objectif 3

- dans la structure de la notice inclure le résumé et le lien direct ou indirect vers le document.

Les points importants pour la mission a traiter:

- Etude de documents existants
- Etude de la chaîne documentaire

Etude et choix: réalisation à niveau de maquette.

Pour l'objectif 3: réalisation du catalogue (avec les notices).

Les autres points demandent une légère étude.

## **VIII.- ETUDE RELATIF A L'UTILISATION DU SERVEUR**

### **8.1.- Statistiques :**

Cette étape a consisté en étudier et analyser le niveau d'utilisation du serveur Handy à travers des applications informatiques disponibles en Internet.

#### **8.1.1.- W USAGE 3.2 :**

Ce logiciel donne une série de fonctionnalités pour connaître l'utilisation d'un serveur. Il donne une variété d'aspects statistiques essentielles pour évaluer l'usage d'un service d'information en ligne. ( Voir Annexe WUSAGE 3.2).

#### **Information obtenue:**

Usage par semaine: Total des accès, total des indexes accèdes, nom des postes connectés. Elle donne, le nombre de renseignements faits dans la période, le nombre total de postes connectés

dans la période, le nombre de questions par HTML, le nombre de questions avec erreurs, le nombre total de questions, le pourcentage de questions par jour et par heure.

### 8.1.2.- GETSTATS , version 1.2:

Nous avons choisi d'approfondir l'analyse statistique avec Getstats version 1.2; analyseur de serveurs WWW capable de prendre le fichier log pour CERN, NCSA, Plexus, GN, MacHTTP ou serveurs Gophers sur Unix , pour faire des statistiques. Cette application nous donne des informations telles que : rapport par mois, rapport par semaine, rapport par jour, sommaire par jour, sommaire par heure, rapport par heure, rapport des questions faites, rapport par domaine, repertoire des rapports.

<b>Rapport par Mois</b>	Couverture de la période en étude	Nom du Mois	Nombre de questions
<b>Rapport par semaine</b>	Couverture de la période en étude	Couverture de chaque semaine	Nombre de questions
<b>Rapport par journée</b>	Couverture de la période en étude	Jour spécifique	Nombre de questions
<b>Sommaire par Jour</b>	Couverture de la période en étude	Nom du jour	Nombre de questions par jour (ex:total pour les Lundis)
<b>Sommaire par heure</b>	Couverture de la période en étude	Heure (ex: 1:00 am)	Nombre de questions
<b>Rapport par heure</b>	couverture de la période en étude	Jour et heure d'utilisation réelle	Nombre de questions
<b>Statistiques complètes du serveur</b>	Couverture de la période en étude en ordre décroissant	a) Nombre de questions	date, Nom du poste connecté
		b) Nombre de questions	date, type de fichier consulté
<b>Statistique par domaine</b>	Couverture de la période en étude (jour en ordre décroissant de connexion)	Nombre de questions	date de dernier connexion, nom de domaine ex: fr. , ca , cl, ch, etc.
<b>Statistique par repertoire ou fichier</b>	Couverture de la période en étude	Nombre de questions	Date de dernier accès et nom de repertoire ou fichier. ex: /sfot, /arch, / : fiche-pascal, etc.

### 8.1.3.- Statistiques sous WWW. Stats 1.0

Le programme wwwstat traite le fichier access\_log du daemon HTTPD de NCSA et produit en sortie un fichier HTML prêt à être affiché sur le Web. Ce programme a été écrit en langage PERL et une fois installé sur votre site, doit travailler sur un système UNIX avec Perl 4.019 ou bien une version plus récente. Il a été basé sur un logiciel des statistiques appelé FWGSTATS -0.035 (par Jonathan Magid) pour multiserveurs.

#### Information obtenue:

- a) Statistiques de transmission par journée
- b) Statistiques par heure
- c) Total de transferts par domaine -client
- d) Total de transferts par sous domaine-client
- e) Total de transferts depuis chaque section ou fichier
- f) Sommaire de statistiques prévues pour la période complète

<b>1</b>	% questions	% bytes	bytes envoyés	n° questions	date (m/j/a)
<b>2</b>	% questions	% bytes	bytes envoyés	temps	-----
<b>3</b>	% questions	% bytes	bytes envoyés	n° questions	domaine (par pays)
<b>4</b>	% questions	% bytes	bytes envoyés	n° questions	sous domaine (par poste)
<b>5</b>	% questions	% bytes	bytes envoyés	n°questions	section ou fichier

### 8.1.4.- Base de Données Access pour Statistiques:

On a fait l'installation du logiciel ACCESS 2 de Microsoft, avec l'objectif de gérer une base de données avec les informations obtenues à travers du logiciel espion. Les données à traiter dans la base de données ont été prises d'une base de données déjà créée pour VBSTATS et son actualisation est automatique semaine à semaine.

L'avantage de ce logiciel est qu'on peut créer et modifier une série de tables, établir les relations entre les objets définis et réaliser des requêtes en langage SQL pour Access 2 de Microsoft, d'accord aux besoins spécifiques de la Mission Handicap.

**Définition du Logiciel Access 2<sup>5</sup> :** Access est une base de données relationnelle de Microsoft. Elle fonctionne sous Windows et offre ainsi une grande souplesse de travail. Grâce à ses facultés très avancées, elle convient tant pour les applications de loisir que pour un usage professionnel.

**Base de Données:** Une base de données relationnelle se compose de plusieurs tables liées entre elles par des relations multiples. Pas une cellule mémoire n'est gaspillée inutilement car les données intervenant dans différents endroits n'ont pas besoin d'être sauvegardées plusieurs fois. Les avantages qui en résultent peuvent être appréciés, entre autres, lors des modifications puisqu'il suffit de mettre à jour un seul élément pour que l'action se répercute dans les divers endroits concernés.

**Les tables:** Toute base de données comporte une ou plusieurs tables dans lesquelles sont stockées les données. Chaque table se compose d'un ou plusieurs enregistrements qui, à leur tour, sont une collection de plusieurs champs. Les divers champs d'une table peuvent contenir des données de types variés (un champ nécessite obligatoirement un nom, un type et des propriétés). ( Voir Annexe Tables pour Base de Données Access)

Access autorise également à définir la clé primaire. dans une base composée d'un ou de plusieurs champs, une clé primaire sert à identifier clairement chaque enregistrement. La clé primaire n'est pas à définir obligatoirement dans toutes les tables. Elle s'avère indispensable pour créer des relations entre les tables ou pour mettre à jour, en une seule passe, plusieurs tables reliées entre elles. Lors, de la définition de une clé primaire, Access crée automatiquement un index qui aide à inspecter rapidement une table. une seule clé primaire est admise à l'intérieur d'une table.

**Création des relations standards entre les tables:** Il faut créer des relations standards entre les tables après avoir défini les tables et avant de générer des requêtes, des états, etc. Ces relations permettent à Access de fusionner les données de deux tables, par exemple pour les besoins d'une requête. pour créer une relation, il faut toujours que l'une de deux tables possède une clé primaire. ( Voir annexe Relations de la Base de Données)

**Les requêtes:** Une requête permet de sélectionner les données souhaitées parmi une collection d'informations. Ces données remplissent un critère préalablement spécifique, par exemple une requête permet d'extraire, à partir du fichier des sites, tous les postes connectés au serveur. Une requête aide également à trier rapidement des données. Dans la plupart des cas, une table constitue la base de création d'une requête. mais on peut aussi construire une requête des résultats fournies par une autre requête. Il existe de nombreux types de requêtes, la plus utilisée étant certainement la requête Sélection. (Voir Annexe Requêtes définies)

**Le formulaire:** Les formulaires permettent d'améliorer la présentation des données d'une table ou d'une requête. Les données affichées dans un formulaire sont stockées en partie dans la table ou la requête. Le formulaire contient lui-même des informations. (Voir Annexe Graphiques et Tableaux)

---

<sup>5</sup> MICROSOFT. Le grand livre de la programmation ACCESS 2. Paris: Micro Application. 1994. 1181 p.

On a choisi deux types des formulaires:

**Le formulaire Tabulaire:** Ou les valeurs des enregistrements sont rangées dans des lignes et des colonnes. Cela permet d'afficher simultanément de nombreux enregistrements.

**Le formulaire Graphique:** Il peut être représenté de différentes manières, il peut prendre l'aspect d'un graphique en aires, à secteurs ou d'un histogramme.

**Instructions SQL:** Lors de la création de requêtes, Access génère un code SQL correspondant à la requête. SQL est un langage<sup>6</sup> de requête structuré pour les bases de données développé surtout sur les gros et mini-ordinateurs.

Chaque requête est convertie en code SQL avec la commande SELECT. dans sa forme élémentaire, elle se compose des clauses SELECT et FROM. SELECT permet de sélectionner les champs à afficher et FROM mentionne le (les) table (s) nécessaires. Il faut utiliser SELECT pour afficher tous les champs dans la feuille de réponse.

Un Critère dans la requête correspond à la clause WHERE en SQL. Cette clause a pour but de rechercher dans la table de données répondant au critère spécifié. Des conditions multiples sont combinées par l'opérateur AND et l'opérateur OR. La clause LIKE définit un modèle de caractère (filtre) à rechercher dans la table. Ce modèle peut être constitué de caractères simples et de jokers. Il existe d'autres instructions qui correspondent aux opérateurs dans les expressions des critères.

L'expression ORDER BY définit l'ordre des colonnes dans la feuille de réponse en tenant compte des indications fournies dans la ligne Tri de la requête interactive.

## 8.2.- Enquête:

Une autre façon de connaître le niveau d'utilisation du serveur Handinet, est à travers une enquête dirigée vers les utilisateurs du service.

### 8.2.1.- Recommandations Générales:

**8.2.1.1.- Type d'enquête:** On doit appliquer une enquête d'opinion, c'est-à-dire auxquelle enquête où on joindra les niveaux d'information, d'attente, etc. (données subjectives).

**8.2.1.2.- Méthode d'enquête:** Parmi les méthodes d'enquêtes existants, le plus pertinent c'est le questionnaire; il doit se présenter comme un document sur lequel sont notées les réponses ou réactions d'un sujet déterminé.

---

<sup>6</sup> Le code SQL d'Access ressemble à la Norme SQL ANSI-86 Level 1 sans pour autant être parfaitement identique. Il existe de grandes différences eu égard aux types de données dans la mesure où de nouveaux types viennent s'ajouter dans ACCESS alors que certains types ANSI ne sont pas reconnus. L'interprétation de certaines instructions montre également des disparités. De plus, ACCESS utilise le point d'interrogation et l'astérisque en guise de jokers et non le trait de soulignement ( ) et le signe % comme dans ANSI SQL.



### 8.3.- Types des questionnaires:

- a) D'administration directe: le sujet note lui-même ses réponses sur le questionnaire.
- b) D'administration indirecte : l'enquêteur note les réponses qui lui fournit le sujet. Ce type de questionnaire a une série d'avantages à considérer, lesquelles sont applicables à le cadre de la Mission Handicap.

- Permet de contacter un public non immediate , dans le cas de la Mission Handicap on doit contacter des usagers distantes à travers du serveur. Cette connection en ligne permettra une liaison directe avec les utilisateurs enquêtés et la tâche a réaliser sera plus efficace, car le service ne dispose pas des enquêteurs.

- L'on est certain que les personnes à interroger sont capables de s'exprimer correctement à l'aide d'un questionnaire.

- Le questionnaire est susceptible de susciter l'intérêt des enquêtés.

### 8.4.- Administration du questionnaire:

- a) Par correspondance: Les questionnaires doivent être envoyés par courrier électronique aux sujets de l'échantillon choisi, avec prière de répondre de la façon plus complète au document envoyé.

- Faiblesse: Il y a un pourcentage de risque de non réponses, ce qui fausse la représentativité de l'échantillon.

### 8.5.- Etapes à considérer dans l'enquête par questionnaire:

- a) Définir l'objet de l'enquête
- b) Inventaire des moyens matériels mis à la disposition des réalisateurs de l'enquête, exemple: temps, personnel, etc.
- c) Recherches préalables: recherche de la documentation sur le sujet, autres enquêtes déjà faites, etc.
- d) Détermination des objectifs et des hypothèses de travail.
- e) Détermination de la population ou univers de l'enquête : à quelle population les sujets à interroger doivent-ils appartenir.
- f) Construction de l'échantillon: procédé de sondage à utiliser, taille de l'échantillon, etc.
- g) Rédaction du projet de questionnaire, choix de méthode d'administration, choix des questions; mise en forme etc.
- h) Mise à l'épreuve du projet de questionnaire: Pre-test ou enquête-pilote.

i) Réalisation matérielle de l'enquête: Collecte des questionnaires remplis (révision courrier électronique); contrôle des questionnaires.

j) Codifications des questionnaires : élaboration du codes.

k) Dépouillement des questionnaires: plan de dépouillement; choix du mode de traitement, etc.

l) Analyse des résultats: choix des méthodes statistiques; sélection des tableaux de base, etc.

m) Rédaction du rapport.

**\*\* Ces étapes ne s'agitent pas nécessairement d'une rigoureuse séquence chronologique.**

**\*\* Certains peuvent se dérouler de manière concomitante.**

## **IX.- ETUDE CONCERNANT AU BALISAGE HTML**

### **9.1. Dysfonctionnement et Nature du Problème:**

On a étudié les processus concernant à la collecte d'information, la chaîne documentaire et les procédures associées à la réalisation de la gestion de documents électroniques.

#### **Les conclusion sont:**

**a)** Par rapport à la collecte d'information, il n'existe pas une série de procédures systématiques pour réaliser ce processus. Dans ce cadre, il faut contacter les services et filiales pertinentes existantes à l'Université Claude Bernard, pour définir un ensemble de politiques et objectifs orientés à la collecte de documents scientifiques et techniques en français.

**b)** L'étude de la chaîne documentaire a démontré que la procédure de sélection est asystématique, situation qui peut avoir incidence en la prise de décision relative au documents à inclure dans la base de données, c'est à dire, baliser de textes non pertinents par rapport aux besoins des utilisateurs principaux du serveur handy (élèves Handicapés). Malgré ce problème, la Mission a une petite collection de documents scientifiques et techniques de type Cours, mais le problème principal est de résoudre des aspects techniques associés aux procédures de balisage HTML.

## Objetif de cette étape :

### - Améliorer les procédures relatives au balisage HTML .

On a détecté que les solutions doivent être orientées à améliorer la procédure de conversion des fichiers format wp, word, rtf, etc. à fichiers HTML. Cette solution doit permettre aussi, de résoudre la **problématique de reconnaissance des formules mathématiques et chimiques**.

En ce qui concerne à la tâche même du balisage, il faut trouver un convertisseur à documents HTML, qui réalise la procédure de marquage et **balisage en moins temps** (parce que la Mission ne dispose pas de personnel suffisant).

### Les caractéristiques du convertisseur à trouver sont:

- Conversion des fichiers formats wp, word, rtf, etc à fichier HTML.
- Reconnaissance des images, tableaux et formules mathématiques et chimiques.
- Convertisseur compatible avec l'infrastructure technologique de la Mission.
- Convertisseur facile à gérer et qui permet de faire le balisage en moins temps, de façon automatique.

## 9.2.- Tâches réalisées :

### 9.2.1.- Delimitation des besoins et Recherche en Internet:

Pour arriver à obtenir la solution, on a décidé de chercher dans la vieille technologie. En ce qui concerne aux activités réalisées dans cette étape, on a commencé pour définir les besoins techniques et caractéristiques du convertisseur à chercher, pour après réaliser la recherche thématique en ligne, dans les différents serveurs et sources d'information en ligne disponibles en Internet.

### 9.2.2.- Etude des résultats obtenus :

Tous les résultats obtenus de la recherche ont été étudiés entre le responsable de la Mission, l'analyste programmeur et la stagiaire, afin de trouver un outil informatique pertinente qui prend en compte des aspects techniques et budgétaires.

### 9.2.3.- Sélection des convertisseurs:

Par rapport aux résultats obtenus dans le cadre théorique, on a choisi d'approfondir sur 4 outils qui sont d'accord à nos besoins; ils sont:

#### a) Cyberleaf: (Voir Annexe Convertisseur Cyberleaf)

Le logiciel Cyberleaf permet de créer des documents composés de texte ou de graphiques hyperliés et de les préparer à être lancés sur le réseau Internet, à partir des documents créés avec des applications de traitement de texte courantes. Cyberleaf peut également relier de multiples documents en un <<Web>>.

Les documents et *Webs* (toiles) qui ont été traités sont au format du langage HTML (HyperText Markup Language) et GIF (Graphic Interchange format), les standards *de facto* d'Internet et du World Wide Web. Ainsi, ils peuvent être visualisés par n'importe quel browser standard tel que Mosaic.

Cyberleaf est une application autonome multiplate-forme qui s'adaptera aisement à votre environnement existant de production de documents. Il reconnaît divers formats de fichiers texte et graphique, ce qui permet de continuer à travailler avec des outils autres de prédilection.

### **Fonctions clés de Cyberleaf:**

- ® **Interface utilisateur graphique:** Une simple interface utilisateur orientée déroulement des tâches guide aisement pendant la conversion au HTML, l'édition d'hyperliens et le processus de création de Webs. Il faut pas connaître la syntaxe HTML ou les concepts SGML.
- ® **Conversion de multiples formats:** Les entreprises peuvent utiliser une large gamme d'outils auteurs destinés à la création des documents: tout sera standardisé par un produit de conversion HTML sur une large palette de plates-formes matérielles et de systèmes d'exploitation. Elles n'ont pas besoin d'intégrer des progiciels auteurs propres à HTML dans leurs systèmes.
- ® **Conversion simultanée du texte et des graphiques en HTML et GIF:** Des documents complexes, incluant texte et graphiques, peuvent être convertis en monopasse, d'où une productivité maximale. Cyberleaf convertit les documents Inteleaf, frameMaker, Ms Word, word Perfect et ASCCI en HTML et les graphiques CGM, IGES et HP-GL en GIF.
- ® **Edition d'Hyperliens facilitée:** L'édition d'hyperliens entre HTML, GIF, PostScript, MPEG, JPEG, le son, l'infographie par quadrillage et d'autres fichiers permet de créer des Webs multimédia pleinement fonctionnels.
- ® **Respect des styles :** La correspondance entre les styles de documents, leurs composants et leurs équivalents HTML élimine le remaniement coûteux des documents ou le développement de masques personnalisés.
- ® **Maintien des hyperliens natifs :** Le transfert des hyperliens des fichiers source natifs FrameMaker, HyperLeaf et WinHelp accélère la conversion de documents hypertexte existants, tels que des logiciels ou des manuels de maintenance au format HTML, sans qu'il soit nécessaire de relier les documents.
- ® **Modèles des pages d'accueil :** Des exemples de pages d'accueil conçus avec professionnalisme sont inclus, comprenant texte et graphiques. Ils peuvent être utilisés pour les applications de distribution des informations courantes, telles que les informations de l'entreprise sur les produits ou le marketing, l'assistance client, les ressources humaines et la formation.
- ® **Mises à jour incrémentales :** Lorsque les documents source changent, les critères utilisés pour créer des hyperliens entre eux peuvent être réappliqués, ce qui permet des mises à jour incrémentales de grands Webs et de documents provenant d'une source unique.

## Prix et disponibilité de Cyberleaf:

Le prix recommande de la version Windows de Cyberleaf est de 3750 FHT et de 5970 pour la version UNIX; il est disponible sur plateformes Digital OSF/1 AXP, HP 700/8xx, IBM RS/6000 et Sun Sparc (O&()). La version PC tournant sur Windows 3.1 /3.11 et Windows NT sera disponible courant 1995.

### b) RTFtohtml: (Voir Annexe Convertisseur RTFtohtml)

Est un filtre qui fait la conversion des fichiers RichTex Format de Microsoft à fichiers HTML. La version plus récente dans le marché est pour supports Word 6.0 Fichiers RTF. Il est utilisé aussi pour fichiers MS-Words ou le format RTF est le document base; il est applicable sur Macintosh, PC ou Unix.

<ftp://ftp.cray.com/Src/wwwstuff/RTF/rtfhtml-overview.html>

page pour version 2.6 : <http://www.w3.org/hypertext/www/tools/rtfhtml-2.6.html>

### c) Megadots : (Voir Annexe Convertisseur MEGADOTS)

Est un traducteur braille et processeur de texte de DOT computing. la nouvelle version de Megadots 1.5 est capable de lire et écrire fichiers HTML en Word Wide Wed. Megadots import et export presque 100 différents types de fichiers comme: fichiers ascii, des fichiers Macintosh, des fichiers windows, des fichiers words, entre autres à fichiers HTML.

<http://www.well.com/www/user-id>

<http://www.well.com/www/dnavy>

### d) Dotsplus : ( Voir Annexe DotsPlus)

Est un méthode experimental d'impression de littérature pour lecteurs aveugles. Il incorpore braille et graphiques et il est recommandé pour l'édition des formules mathématiques en braille.

<http://dots.physics.orts.edu/dotsplus.html>

### e) Autres résultats:

Un des résultats importants de la recherche , est l'information obtenue par rapport à la nouvelle version de HTML 3.0 qui se trouve en développement et si bien elle n'est pas totalement normalisée, permettra de travailler avec formules mathématiques. (Voir Annexe HTML 3.0).

### Description:

La marque <MATH> est utilisé pour inclure une expression mathématique sur la ligne courante. Les marques mathématiques d'HTML sont suffisamment puissante pour décrire le rang des expressions mathématiques que vous pouvez créer dans un traitement de texte commun, pour pouvoir être rendu sur synthèses vocales. Quand l'affichage est seulement fixé

au texte, de simples caractères graphiques peut être utilisés pour les symboles mathématiques comme le signes d'intégration, tandis que d'autres symboles peuvent être affichés par leurs noms d'entités. La capacité SMGL SHORTREF est utilisé pour produire des abréviations pour les accolades cachés les exposants et les indices.

#### **9.2.4.- Essai du convertisseur choisi:**

Après d'étudier les caractéristiques, les possibilités, les faiblaisses et les avantages de chaque convertisseur ou filtre on a pris la décision de solliciter une version de démonstration de Cyberleaf, avec l'objectif de tester son fonctionnement:

## **X.- CREATION D'UN CATALOGUE EN LIGNE POUR LES DOCUMENTS ELECTRONIQUES DU SERVEUR HANDY**

### **10.1.- Etude du Logiciel Doris/Loris: (Voir Annexe Documentation Doris)**

En ce qui concerne au sujet de ce stage, on a étudié la documentation de Doris fourni pour la Société EVER et aussi l'information et caractéristiques de LORIS et OPAC (information fournie par un spécialiste à l'Enssib); avec l'objectif de connaître ses applications et apercevoir sa mise en place à la Mission Handicap.

Les besoins spécifiques de la Mission sont de créer un catalogue en ligne des documents scientifiques et techniques et mettre à disposition des utilisateurs le document hypertexte à travers une liaison directe avec un champ de la notice bibliographique.

Antecedents:

**Doris/Loris/Opac:** C'est un ensemble pour stocker et gérer grandes masses d'informations.

**Doris:** (Documentary-Oriented Relational InformationSystem) est une synthèse intégrée des méthodes d'organisation et des sciences de l'information. Ce logiciel permet de décrire une réalité complexe et s'adapte toute à fait au stockage des grandes masses d'information de nature très varié (document numérique, document éditeur ou traitement de texte, fax, télex, image analogique, image video, document sonore et document graphique).

**Loris :** permet de gérer les fonctions d'un bibliothèque en forme intégrée, comme sélection, acquisition, catalogage, etc, à travers la définitions des objets, définition des administrateurs, définitions des utilisateurs, définitions des procédures à informatiser, etc.

**OPAC:** C'est l'interface qui permet aux utilisateurs du système de connaître l'état de la collection.

Après de réaliser cet étude on a décidé que l'ensemble Doris/Loris/Opac n'est pas applicable à la Mission parce que:

- Le fond de documents électronique à gérer est petite (10 documents de type CORS et TD)
- La seule fonction à gérer de la Mission est la de catalogage, pour tant l'exploitation de Loris sera partielle.
- Pour l'installation de cet ensemble, est nécessaire l'installation d ' ORACLE, qui permet le stockage organisé de données ainsi que leur manipulation. Il constitue aussi un système de gestion de base de données relationnelle.
- Doris/Loris ne permet pas la liaison directe de la notice bibliographique et le document hypertexte .

## **10.2.- Etude de la collection existante de documents scientifiques et techniques type cours et TD:**

On a étudié la collection existante des documents de type cours et TD disponibles dans le serveur, les résultats de cet étude sont:

- Un fonds de 10 documents techniques et scientifiques d'accès limité à la Mission Handicap (non accessibles aux utilisateurs).

- La structure des documents n'est pas normalisée, on trouve des textes sans éléments de base pour réaliser la description bibliographique, comme par exemple date, éditeur intellectuelle, etc.

L'inexistence de ces types de données aura incidence directe en les critères de recherche et documents récupérés par les utilisateurs. Dans ce cadre, el faut donner de recommandations aux enseignants et remplir la source original avant le balisage HTML, avec les informations et données nécessaires pour réaliser le catalogage d'accord à normes internationales. (Description bibliographique d'accord à l'ISO et AFNOR).

### **Avantages d'une description normalisée:**

- Identification inéquivoque de la source original.
- Exploitation de toutes les données potentielles pour la posterieure récupération d'information sous critères de complétude et pertinence.
- Interchange de notices bibliographiques.

## **10.3.- Etude des données bibliographiques et définition des champs:**

### **10.3.1.- Principe :**

La description bibliographique des sources d'information doit être basée sur normes internationales existantes (ISO, RCAA2) , avec l'objectif de faire une description inéquivoque du chaque document et faciliter la récupération d'information.

Les normes internationales proposent une série de zones , par rapport a chaque type de donnée. En ce qui concerne à la Mission Handicap, on a étudié les normes, les besoins de la Mission , les caractéristiques des documents au sein de la entité et le niveau de connaissance dans le domaine du Catalogue des Collections par le personnel de la Mission.

### **10.3.2.- Structure des Champs de Chaque Notice Bibliographique** ( D'accord ... Norme ISO et RCAA2)

Cette structure a été étudié par le stagiaire avec le responsable de la Mission, l'ingénieur et l'analyste programmeur qui réalise la gestion des documents électroniques. Les résultats de la fiche ou notice bibliographique ont été modifiés, mais sont basés sur les champs suivantes:

#### **a) VEDETTE PRINCIPALE:**

- La vedette principale c'est l'auteur ,s'il y a 1 ou plusieurs dans le document (jusqu' à trois). On doit écrire le premier auteur enregistré dans la source principale.

- La vedette principale sera le titre en absence de(s) auteur (s) enregistrés dans la source principale. S'il y a plus de trois auteurs, la vedette sera aussi par titre.

#### **b) TITRE UNIFORME:**

- C'est le titre général du document .Il doit être écrit s'il a une différence avec le titre propre de la source.

#### **c) TITRE PROPRE et SOUS TITRE:**

- C'est le titre plus détaillé et complet du document. Il est un champ obligatoire. A continuation de celui-ci on doit mettre d'autres informations sur le titre.

#### **d) AUTEUR:**

- C'est un champ obligatoire s'il est mentionne dans la source principale.



- S'il y a plusieurs auteurs, on les doit écrire en ordre d'apparition dans le document.

- S'il y a plus de trois auteurs, on doit écrire dans cette zone de responsabilité intellectuelle ; le premier auteur mentionné dans la source au celui-ci que la personne qui fait le catalogage considère plus important ou relevant; les autres seraient mentionnées avec l'abréviation et. al. ou : et autres.

- Il faut considérer qu'une entité peut être considérée comme un auteur aussi, et celle-la peut être à la fois l'entité responsable de l'édition.

- La description de la responsabilité intellectuelle , peut être présente dans un champ répétitif, ou à travers la création des champs Auteur 1, auteur 2, auteur 3.

#### e) TRADUCTEUR:

- C'est un champ de caractère obligatoire, si le document est une traduction d'une autre langue de la source originelle.

- Si le document c'est une traduction, il faut faire dans la zone de notes, la note relative au titre original, langue et toutes les informations pertinents pour arriver à la source original

#### f) LIEU D'EDITION :

- Pour la description d'un document, on considère le pays ou ville, comme le lieu de l'édition ou production de la source.

- Dans le cas des documents électroniques du serveur de la Mission Handicap , ce champ doit être de caractère fixe ou prédéterminé,. Ex: Villeurbanne ou France ou Villeurbanne, France.

#### g) EDITEUR/ PRODUCTEUR:

- C'est un champ de caractère prédéterminé ou fixe. Pour les documents de la Mission sera toujours l'Université Claude Bernard, Lyon I.

- Pour compléter l'information de cette zone, on peut créer un autre champ Sous Editeur ou créer un sous champ, avec l'information relative au laboratoire, département, section de l'université; responsables aussi de la production de la source.

- La description de toutes les entités relatives à l'édition ou production devient être enregistrées d'accord à un critère hiérarchique.

#### h) DATE:

- C'est un champ de caractère obligatoire, on doit enregistrer dans cette zone la date ou année d'édition ou production du document; Si elle n'est pas enregistrée dans le document, on doit écrire s.d. (ou l'abréviation proposée par l'AFNOR).

## i) DATE de M.A.J.:

- C'est un champ qui sera rempli de façon automatique par le système.
- La structure de présentation de cet information, doit être prédéterminé, Ex: AA/MM/JJ ou JJ/MM/AA .

## j) MOTS CLES :

- Les mots sont les matières données au documents. Chaque mot clé représente le contenu spécifique du contenu de la source, elle peut être un champ réapétitif. En plus, elle doit être créée par la personne qui fait le catalogage, d'accord à sources spécialisées de langage contrôlé (thesaurus), ou d'accord à une liste locale de matières de la Mission.

## k) NOTES:

- Cette zone est relative à toute l'information relevante que complète la description de la source . Elle est divisée en plusieurs sections (9 zones d'accord à l'ISO), sont optatives et font relation a notes associées à :

- Note de responsabilité intellectuelle du document
- Note relative au titre du document
- Note de traduction: ex. titre original, auteur, etc.
- Note de contenu: ex. parties ou chapitres du document.
- Note de propriété intellectuelle
- Note pour les utilisateurs : ex. autres versions, couverture, accès, etc.
- Autres

## l) RESUME:

- Le résumé est une outil pour représenter le document complet dans une version abrégée .
- L'extension de ce champ peut être à deux niveaux dépendant du type de résumé:

a) Indicatif jusqu' à 250 caractères.

b) Explicatif jusqu' à 450 caractères.

- Dans la réalisation du résumé, il faut considérer les mots clés pour la recherche documentaire posterieure.

## m) LANGUAGE:

## n) ADRESSE ELECTRONIQUE:

- Champ destiné aux adresses ,électroniques à considérer pour la liaison hypertexte en HTML.

#### **10.4.- Visualisation des futurs problèmes de recherche:**

On a étudié possibles problèmes relatifs àaux critères de recherche bibliographique du coté des utilisateurs. Eux sont associés à la pertinence de l'information obtenue, complétude. Pour éviter cette problématique on trouve plusieurs sources spécialisées dans le domaine de l'indexation ,bibliothéconomie et documentation.

Ces thésaurus permettent de normaliser les descripteurs ou matières représentatives du contenu de la source en description. Le problème à résoudre pour la Mission est alors de normaliser les mots clés avec l'aide d'une outil facile à gérer par personnel non spécialisée en cette tâche. dans ce cadre on a éliminée la possibilité d'un thésaurus car il présente des différents relations entre les concepts (des relations associatives, hiérarchiques,etc). La solution temporaire est de créer un liste locales des matières modifiables d'accord aux besoins de la Mission, mais toujours sous un critère de Mots normalisés.

La liste suivante est proposée pour résoudre cette problématique a court terme, mais est nécessaire faire un étude plus approfondi. Elle est basée dans las matières en etude dans l'Université Claude Bernard- Lyon I. (Voir annexe Matières en étude a l'Université Claude Bernard )

#### **MATIERES EN ETUDE A L'UNIVERSITE CLAUDE BERNARD**

##### **A**

Aaudioprothèse

##### **B**

Biochimie

Biologie

Biologie appliquée

Biologie humaine

Biologie medical

##### **CH**

Chimie

##### **C**

Commerce

##### **E**

Ecologie

Electronique

Electrotechnique

Ergothérapie

**G**

Génie medical  
Génie biologique  
Génie biomédical  
Génie Civil  
Génie électrique  
Génie mécanique  
Gestion  
Gestion de la qualité

**I**

Image et productivité  
Images des Synthèses  
Industrie chimique  
Industrie du médicament  
Industrie hospitalière  
Informatique  
Informatique Documentaire  
Ingénierie mathématique

**K**

kynésithérapie

**M**

Maintenance industrielle  
Mathématiques  
Mécanique  
Mécanique énergétique  
Médecine  
Microbiologie  
Mulitmateriaux

**O**

Odontologie  
Orthopedie

**P**

Pharmacie  
Pharmacie industrielle  
Physiologie  
Physique  
Productique et mathématiques appliqués  
Psychomotricité

**S**

Santé

Sciences

Sciences appliquées

Sciences de la terre

Sciences de la vie

Sciences pour l'ingénierie

Secteur tertiaire

**T**

Techniques de réadaptation

Technologie industrielle

**10.5.- Maquette de fiche bibliographique:****Titre (ti) :** Almanach anti-superstitieux**Auteur (au) :** CONDORCET**Lien vers le document document html****Document RTF :** sources rtf**lieu d'édition :** Villeurbanne, France**éditeur :** Anne-Marie CHOUILLET, CNRS EDITIONS, UNIVERSITE DE ST ETIENNE**source (so) :** Laboratoire d'Analyse Numerique**filillère universitaire (fi) :** DEA Litterature 18 eme**date du document source :** 1992 date du document source**date de création du document html :** 95/05/19 date de creation du document html correspondant (annee)**date de mise à jour :** 95/09/06 date de mise a jour du document**type de document :** BI type de document (Cours:CR / Travaux diriges:TD /Document  
www:WW/Bibliotheque:BI)**mots-clefs (kw) :** revolution, religion, superstition**notes divers :** notes, diverses concernant la gestion du document

18 siecle date du document original

ppte intellectuelle, contenu (decoupage),

notes pour les utilisateurs..

**résumé :** abstract

**langue :** Fr

**taille source :** taille du document postscript complet (pages, octets)

**taille document html :** 477309 taille du document html (lien direct) ( octets)

**responsable mise à jour :** Responsable du document [htmlbayard@handy.univ-lyon1.fr](mailto:htmlbayard@handy.univ-lyon1.fr)

**adresse de l'auteur :** contact mission handicap: [crepel@lan1.univ-lyon1.fr](mailto:crepel@lan1.univ-lyon1.fr)

-----  
Script SFgate 3.6 written by Ulrich Pfeifer. Source is available.

## **10.6.- Procédures:**

- a) Création de cartuche dans fichier HTML
- b) Indexation de fichiers pour base de données wais
- c) Essai d'édition de notice bibliographique
- d) Essai de recherche documentaire sous différents champs de la notice bibliographique.

## CONCLUSION

La réalisation de ce stage à constitué une expérience personnelle est professionnelle aussi intéressante qu'enrichissante. Les raisons pour faire une affirmation de ce genre, sont basées sur plusieurs aspects.

L'insertion comme professionnelle étrangère dans une entité française, m'a donnée la possibilité d'apercevoir comme le développement des technologies avancées permettent l'incorporation des utilisateurs au monde de l'information.

Dans ce cadre, il faut diffuser dans les systèmes d'information traditionnels de pays en train de se développer, la notion d'information à tous les utilisateurs, quelque soit leur type

L'accès à une infrastructure technologique avancée permet de mettre à disposition des usagers un grand volume d'information, mais malgré ça, on rencontre des problèmes que comme documentaliste il ne pas facile d'apercevoir; par exemple que l'existence d'un serveur en ligne et l'accès à Internet n'assure pas que les utilisateurs auront l'accès à l'information pertinente. C'est le cas du serveur Handy qui est destiné à un groupe des utilisateurs ayants des besoins bien définies et avec des limitations pour accéder à l'information selon leur incapacité.

En ce qui concerne la Mission, elle réalise une activité très importante et peut constituer dans le futur, une entité clé dans le monde des étudiants handicapés en France.

Mais il ne faut pas oublier que la création d'un système d'information doit passer nécessairement par la définition des objectifs, où toutes les activités à réaliser et les services à offrir doivent être orientés vers un groupe très spécifique des utilisateurs.

Le problème le plus important qu'on a trouvé à la Mission Handicap est associé à la multiplicité des projets à réaliser, situation qui a eu une incidence directe sur les utilisateurs du système, par manque des priorités. Mon expérience administrative comme bibliothécaire et documentaliste me conduise à faire quelques recommandations à ce sujet :

- Il faut définir des objectifs clairs et précises à court, moyen et long terme;
- Il faut définir les catégories des utilisateurs et prendre en compte toutes les activités à développer à partir de la définition adoptée;
- La planification des activités est très importante pour arriver aux objectifs définis a priori.

- Si bien qu'il faut améliorer des aspects techniques de type informatique; il faut aussi faire attention aux procédures associées à la collecte d'information, car le un système d'information tel qu'il est développé à la Mission Handicap risque de conduire à obtenir une collection non pertinente par rapports aux besoins des utilisateurs.

Comme étudiante d' Informatique Documentaire, je conclurais que la formation dans le domaine de l'informatique est vitale pour les spécialistes de l'information. Aujourd'hui, avec le développement des télécommunications, des réseaux et d'en particulier d'Internet, se fait indispensable une connaissance technique, car aucun bibliothécaire ou documentaliste ne peut créer un système d'information automatisé, sans la connaissance approfondie.

Un cas très fréquent dans les bibliothèques et les centres de documentation est de prendre la décision d'automatiser des procédures relatives aux différentes sections du système, sans prendre en compte l'état de la veille technologie.

Dans la réalité on trouve des problèmes techniques difficiles à résoudre, et alors devient nécessaire un travail continu de recherche et de mise à jour dans le monde de l'informatique. Mais à mon avis, l'aspect le plus déterminant à considérer est la réalisation d'une activité interactive entre les différents spécialistes des sciences de l'information. De cette façon, une équipe de travail peut arriver à trouver les solutions pertinentes pour améliorer un système d'information, soit traditionnel , soit très développé comme les serveurs disponibles en ligne à travers Internet.



## I- ANNEXES : PROJETS DE LA MISSION :

### 1.1.- Projet PRINCE:

Son objectif est développer un système d'élaboration et de gestion coopérative de documents multimédia **par intégration de technologies existantes**.

#### ® **Avantages de PRINCE:**

- Le système **PRINCE** apporte aux outils actuels des principes d'architecture et un outil qui permettent d'**organiser toute la production des documents**, dans le cadre d'un projet, en intégrant les principes de qualité de la norme **ISO 9000**.

- Le coeur du système **PRINCE** est une **base de données active**, permettant le déclenchement automatique des règles de production des documents.

- Toutes les informations contextuelles des documents sont intégrées dans le système d'information de l'entreprise, permettant **une réelle gestion des documents**, et par la même **une réelle gestion des projets concernées par ces documents**.

- La **navigation de type hypertexte**, à partir des outils des utilisateurs, devient possible entre les documents et les données du système d'information.

- **PRINCE** bénéficie de tous les apports de la gestion documentaire, notamment la consultation, développés autour de l'outil **DORIS**.

® **Positionnement de Prince:** Prince est un outil Groupware puisqu'il est défini comme une application "d'élaboration et de gestion coopérative de documents multimédias".

#### ® **Fonctionnalités de Prince**

	PRINCE
Agenda électronique de groupe	OUI
Messagerie / courrier électronique	OUI
Workflow	OUI
Editeur partagé	NON+
Rédaction coopérative	OUI
Forum électronique	NON
Visioconférence	NON
Aide à la décision	OUI / NON

**a) Agenda électronique:**

L'agenda électronique devra mettre à jour les échéances individuelles de rédaction. Il est utilisé pour la planification des tâches des auteurs du document commun.

**b) La messagerie**

La messagerie permet la communication entre les acteurs d'un même travail coopératif

**c) Workflow:**

Le workflow sera utilisé pour gérer la circulation des documents et des différentes versions entre les acteurs. Flow Path a été choisi comme moteur de workflow.

**d) Rédaction coopérative:**

Ce système de rédaction coopérative sera asynchrone au quasi synchrone. Il utilise word 6 comme outil de base. PRICE améliore les performances de word 6 en y greffant un système de workflow permettant une gestion des tâches des auteurs des documents, en améliorant le système d'annotation et de révision des documents et en permettant de visualiser l'historique des modifications.

**e) Aide à la décision:**

Le système d'aide à la décision n'est pas compris dans le sens des outils d'aide à la décision du marché. Il permet au rédacteur en chef d'intégrer les commentaires des autres acteurs après les avoir traités.

Il est défini comme un système d'aide à l'élaboration de synthèse.

**f) Autres fonctionnalités de PRINCE**

PRINCE permettra la manipulation des documents multimédias en intégrant la technologie OLE de Windows. Grâce à DORIS, PRINCE proposera une gestion documentaire beaucoup plus poussée que Word 6.

**® Spécifications générales du projet PRINCE:**

Le projet **PRINCE** propose d'utiliser des outils quasiment standards dans les entreprises comme **WORD** et **EXCEL**. Il se articule autour de **DORIS**, la base de données de type objet de la Société **EVER**.

L'outil **DORIS** est lui-même une sous-couche de moteurs relationnels courants (**ORACLE**, **SYBASE**, **INFORMIX**). Pour les entreprises, le projet utilisera **DORIS**, un moteur relationnel supporté par **DORIS**, **WORD**, **EXCEL**.

**© Le système PRINCE est composé en 3 sous-systèmes:**

**a) rédaction multi-acteurs de documents:** C'est la gestion des éléments intervenant dans la production du documents multimedia, notamment la gestion des acteurs, des contenus multimédias, des structures, des versions, de la relecture, des annotations multimédias, des règles de production et des échéanciers.

**b) Travail coopératif et aide à la décision:** Le premier système intègre le travail coopératif par le découpage du document en partie pour la rédaction et par le processus de relecture. Ce deuxième sous-système propose d'aller plus loin dans le travail coopératif,

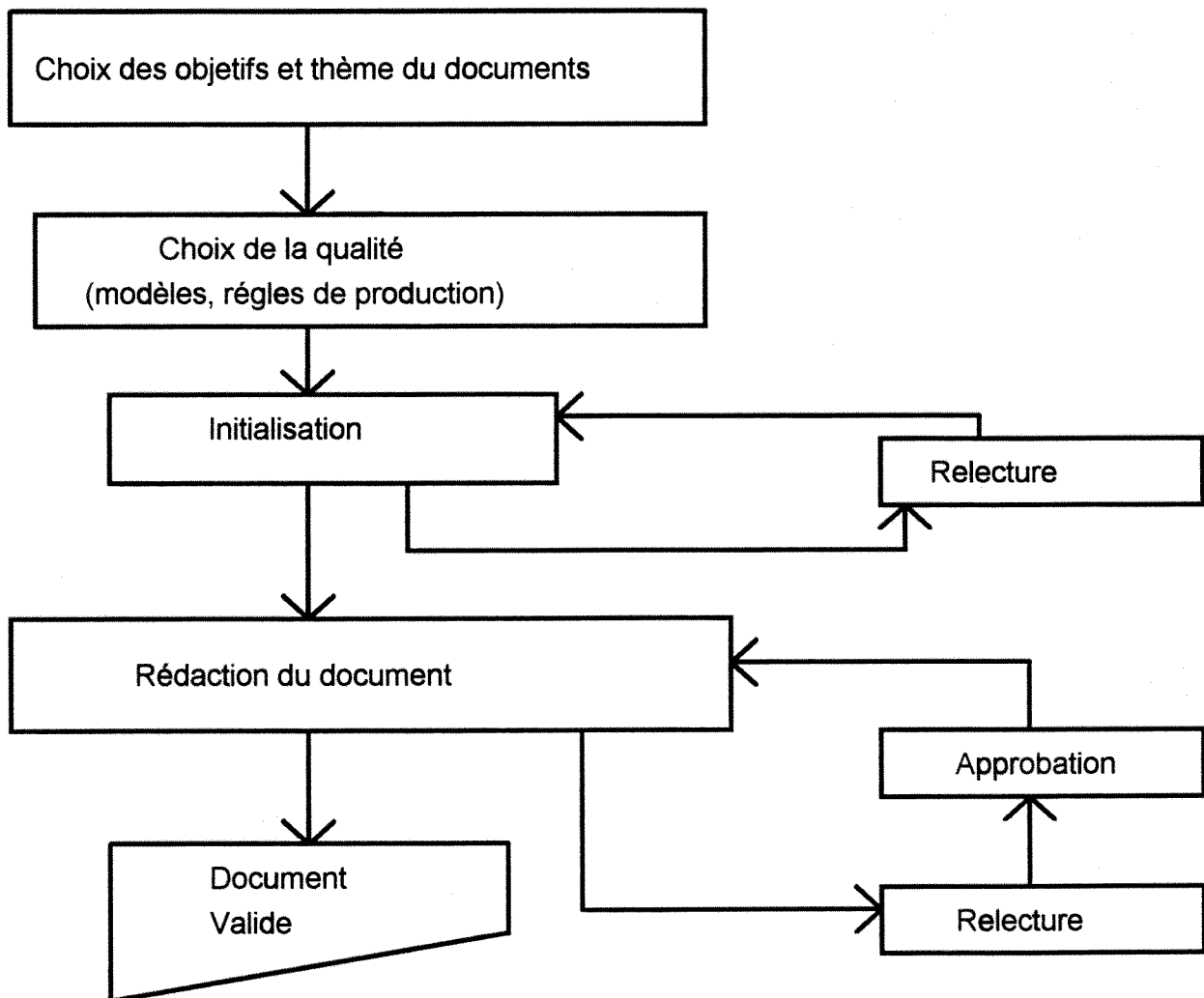
notamment en abordant le travail simultané et interactif de plusieurs personnes autour du même document.

**c) Contrôle de cohérence a posteriori et liens avec le système d'information:**

Il s'agit de définir les interfaces avec les bases de données existantes, et de gérer les liens entre les documents et les bases de données.

La description du système **PRINCE** s'insère dans le cadre du cycle de production qui suit:

**CYCLE DE PRODUCTION**



**®Conclusion:**

**PRINCE** n'est pas particulièrement novateur par rapport aux fonctionnalités proposées car il intègre des outils déjà existants sur le marché. Sa valeur réside dans l'articulation de technologies déjà existantes (workflow, agendas, documents multimédias, messageries, ...) et l'ajout d'un système de gestion documentaire évolué grâce à **DORIS**, technologie développée par **EVER**.

## 1.2.- Projet CRISTALLIN :

Ce projet a pour but de développer un système de gestion et de manipulation des documents électroniques et d'aide à la décision dont l'interface homme-machine sera spécifiquement adaptée aux besoins des déficients visuels (**non voyants**).

En effet, **CRISTALLIN** sera conçu en parallèle et comme une extension de **PRINCE**, dont la construction sera réalisée en tenant compte des spécificités des déficients visuels. De cette façon le projet permettra d'intégrer, au fur et à mesure de la conception de **PRINCE**, les modifications particulières eu égard aux déficients visuels.

Ce projet s'inscrit dans la continuité du projet **PRINCE**, qui vise à développer un système d'élaboration et de gestion coopérative de documents multimedia. Il s'appuie sur le constat suivant : les déficients visuels présentent des aptitudes particulières (mémoire, esprit de synthèse...) qui leur permettraient de remplir aussi bien qu'une personne voyante.

Ainsi **CRISTALLIN** a pour but le développement d'un système de gestion de documents et d'aide à la décision, dont l'interface homme-machine sera spécifiquement étudiée afin que déficients visuels et voyants puissent utiliser confortablement cet outil.

Le projet propose d'utiliser les outils quasiment standard dans les entreprises tels que Word et Excel en s'articulant autour de **DORIS** sur-couche de moteurs relationnels courants comme **ORACLE**.

La modélisation des documents devra être conforme à la normalisation internationale de SGML et capable de représenter de façon intelligente de textes scientifiques, juridiques, etc.

### ®Récupération et structuration SGML de documents

Par ce projet **CRISTALLIN** d'adaptation de **PRINCE** aux déficients visuels, le travail de la Mission Handicap consistera à fournir des documents **SGML** (norme incluant les notations scientifiques) à la base de données de **PRINCE**. D'une part, il faudra rassembler ou créer les outils nécessaires à la production de documents pour la base de données et d'autre part créer cette base et l'alimenter.

**Objetif:** Les fonctionnalités de **PRINCE** (rédaction et accès aux bases de données) sont actuellement accessibles **séparément** aux non voyants. Le projet **CRISTALLIN** vise à **intégrer en un seul système**, à leur donner une **ergonomie maximale tout en proposant une solution à tous les points posant un problème particulier** (lecture de tableaux, images, graphiques...):

- reconception des objets des interfaces graphiques,
- adaptation des objets non-visuels à l'interface homme-machine ; intégration des périphériques,
- construction de l'interface en tenant compte des conclusions des travaux antérieurs et de la sémantique de l'application visée, avec le **souci de définir des concepts**

**suffisamment généraux par que les travaux effectués dans ce cadre soient réutilisables.**

® **Public visé:** Dans un premier temps, les auteurs du projet, conscients des différences entre malvoyances et non-voyance et de la variété des malvoyances, s'intéresseront exclusivement à la cécité.

L'accès à la documentation et aux procédures liées à un projet (simple ou complexe), ainsi que les facilités offertes par la rédaction multi-acteurs permettront aux non-voyants d'assurer **des fonctions administratives, de secrétariat et des gestion des projets**. Grâce à un outil tel que CRISTALLIN, l'insertion professionnelle de non-voyants serait hautement amplifiée.

® **Positionnement de Cristallin:** Crsitallin est défini comme une application permettant aux non-voyants d'accéder à la totalité des fonctionnalités de PRINCE (Production Intelligente et Coopérative de Documents en Entreprises). Grâce à des périphériques "braille" ou "synthèse vocale", le non-voyant pourra avoir accès à l'environnement windows. En plus, grâce à SGML supporté par windows, les non-voyants auront accès à une synthèse des documents produits par les membres du groupe.

#### ® **QUELQUES DIFFERENTIATIONS ENTRE CRISTALLIN ET PRINCE**

- Le projet **PRINCE** est un projet **pré-industriel** avec une vocation **d'industrialisation**.
- Le projet **CRISTALLIN** combine des aspects **recherche** avec une volonté **d'industrialisation**.

#### ® **La faisabilité technique de PRINCE**

Bien que le projet PRINCE n'ait pas fait l'objet d'une étude de faisabilité formelle, l'étude du projet, des moyens humains mobilisés, du potentiel technique, des méthodes des documents de spécifications (spécifications générales et spécifications détaillées) permettent d'affirmer que le projet PRINCE est faisable.

Les arguments sont les suivants:

-**PRINCE** est un projet d'intégration de produits et d'outils existants dans un environnement logiciel et matériel standard.

-Le potentiel technique des deux partenaires permettra de surmonter les problèmes techniques auxquels ils seront confrontés.

-Les enjeux stratégiques pour l'un des partenaires (**la société EVER**) sont une garantie de son implication maximale dans la réussite du projet.

## ®La faisabilité technique du Projet CRISTALLIN

-Le projet **CRISTALLIN** est un projet de recherche.

-Il y a un certain nombre des conditions à la faisabilité de **CRISTALLIN**:

a) L'existence sur le marché d'un certain nombre d'outils logiciels et matériels qui sont susceptibles d'être incorporés dans l'interface homme-machine.

Cependant, les grands éditeurs (Microsoft) ne sont pas inactifs, la prochaine version de windows intégrera un traducteur **SGML**. La norme **SGML** et son respect étant essentiels au projet **CRISTALLIN**.

b) La permanence et la compétence des équipes impliquées dans la réalisation.

Le projet **CRISTALLIN** dispose de quelques atouts:

- L'implication d'un industriel (**EVER**)

- Un encadrement scientifique de haut niveau (**LISI**)

- Une mobilisation de structures connaissant bien les publics concernés, avec une pratique de recherche pour la mission handicap et de formation, conseil pour la **FIDEV**.

- Des acteurs institutionnels très motivés pour la plupart,

**L'accessibilité de PRINCE et CRISTALLIN pour tout public handicapé**

L'accessibilité **CRISTALLIN** à tout public handicapé ne figure pas au cahier des charges de ces 2 projets. On peut penser que l'attention toute particulière qui sera portée à l'interface homme-machine est susceptible de rendre le produit **PRINCE** plus accessible à d'autres handicaps. Les attentes quant à tout public sont limitées. En effet, une interface pensée pour les déficients visuels risque de répondre difficilement aux besoins des déficients auditifs, par exemple.

### 1.3.- PROJET CAPESA :

Ce projet se réalisera avec la Société **CAP GEMINI INNOVATION**, entreprise qui propose la création d'un logiciel pour la fabrication automatique de résumés. La participation de la Mission Handicap dans ce projet est d'apporter sa compétence dans le domaine de la création et gestion de serveur sur réseau de documents électroniques. Les partenaires potentiels sont la Organización Nacional de Ciegos de España (**ONCE**), l'Union Européenne des Aveugles et la Société Française CAP Gemini Innovation.

Les caractéristiques techniques du logiciel destiné à la création automatique de résumés, mais dans un cadre théorique cet outil informatique doit être compatible avec les outils déjà existants en la Mission. Une remarque importante à faire c'est que pour la création d'une éventuelle base documentaire de gestion de catalogue de documents électroniques, il faut considérer un champ destiné au résumé et son postérieure liaison au document en hypertexte.

## II.- ANNEXES : LOGICIELS POUR STATISTIQUES

### 2.1.- WUSAGE 3.2

usage statistics for WWW servers

**IMPORTANT NOTE:** Many people are emailing me saying that wusage creates zillions of files dating back to 1970, et cetera. This happens when you set the wrong value for your server type in wusage.conf. Only OLD versions of the NCSA server should have the NCSA\_HTTPD setting; anything from 1.2 on should have the COMMON setting. Keep this in mind as you read the configuration section. We now return you to your regularly scheduled documentation. -

#### Table of Contents

- \* What's New in version 3.2!
- \* Credits and license terms
- \* What is wusage?
- \* What else do I need to use wusage?
- \* What if I don't use the NCSA server?
- \* How do I get wusage?
- \* How do I build wusage?
- \* Configuring wusage for your server
- \* Excluding unwanted accesses
- \* Charting access by domain
- \* Installing wusage as a cron job
- \* Purging access\_log (how and why)
- \* Please tell us you're running wusage!
- \* If you have problems

#### Up to Quest Home Page

#### What's New in version 3.2!

Support for prefix and suffix files of your own design. These let you completely control what you add to your usage pages, and eliminate the old "Up to Home Page" link. You will need to edit your wusage.conf (required) and can prepare prefix and suffix files of your own (optional)!

Support for VMS. Patches provided by Robin Garner. A bug fix: the domain charts used to be based on all usages, regardless of settings for sites and documents to be ignored. Now they are based only on accesses included in the total access count. (Ignored sites and documents are not included. Hidden documents are included.)

Support for the PLEXUS server log format. Provided by Glenn Heinle. I have not tested this code myself; you can contact him at [heinle@cmf.nrl.navy.mil](mailto:heinle@cmf.nrl.navy.mil) if you have questions about the PLEXUS support.

Bug fix: you can now ignore/hide/etc. accesses of your root document by blocking accesses to `"/"`.

Compatibility enhancement: since the CERN server uses "Welcome.html" and "welcome.html" in the way the NCSA server uses "index.html", all three are now aliased to "/" when used in the root directory.

Uppercase site names are now converted to lowercase before being analyzed.

Identity check information (person@) is now stripped off of site names before they are analyzed.

What's New in version 3.1!

DOMAIN CHARTS! wusage can now create pie charts showing access by domain, and lets you alias country domains together to make statistically significant continent domains.

A bug fix for the double-reports that have appeared recently in 3.0.2. Hopefully. Hard to tell since I've seen them at my site but attempts to deliberately produce them fail. Let me know what happens with 3.1 if you've seen them with 3.0.2.

Compatibility fix: strdup() has been banished, replaced by a simple mystrdup() function. This will make those whose C libraries lack strdup() happy.

I have received patches for VMS and am working on built-in MSDOS support, among other things, for the next release. My apologies to those who hoped to see them in 3.1.

What's New in version 3.0.2!

A small fix: graphs weren't coming out right on the first pass, but were OK if wusage was run a second time. Fixed -- graphs now come out right the first time.

What's New in version 3.0.1!

Just a small fix to version 3.0: graphs were displaying an extra week with zero accesses. Fixed up.

What's New in version 3.0!

Version 3.0 has been extensively rewritten. It is not so much enhanced as made far more stable and very easy to install:

- \* PBMPLUS is no longer required.
- \* Configuration has changed.
- \* The syntax to run wusage has changed.
- \* wusage is now a single executable. No muss, no fuss.
- \* usagegraph and ppmfig are gone; GIF support is built-in now.
- \* The twice-in-one-week bug is gone! You can run wusage as many times as you like now; it no longer trips on itself when you run it twice in one week.
- \* Non-ANSI C compilers are now supported. I still think this is silly in 1994, but many people just have vanilla Sun CC.
- \* NCSA "old" log file format is now supported again, along with common



logfile format (the newest NCSA httpd version use this) and old-style CERN format. There was a bug in 2.5 which locked out this support.

#### What's New in version 2.5!

Support for the new common log file format.. Those of you who have just upgraded to NCSA httpd 1.2 will note that wusage has stopped working. This is because the log file format has changed. The first line of your wusage.conf file now most likely reads either NCSA or CERN; change it to COMMON, if you have the latest version of either server. Note that data in the old format can't be read once you do this, and will be ignored. If you've been running wusage all along, this is not a problem, since you already have reports for all previous weeks; you may need to fudge your results a bit for the one week during which you switched servers.

Please see configuring wusage for your server.

#### What's New in version 2.4!

Incredibly dumb mistake on my part while creating 2.3 which led to problems even worse than those in 2.2 is now fixed. It's amazing how much trouble one line of code can cause. THIS VERSION WORKS, at least on the systems available to me for testing.

#### What's New in version 2.3!

One-line but extremely important bug fix in wusage.c! I deleted a critical line between version 2.1 and 2.2. Mea culpa.

The bullet is no longer inside the anchor in the list of weeks, owing to a problem with at least one client that won't accept this (although I believe it's valid HTML).

#### What's New in version 2.2!

Bug fixes in several places. Sites with less than one week of data should work fine now (of course the graph is rather dull with no data points, wait a week or two); meaningful error messages should appear for missing directories.

A reference to an out-of-date version of pbmplus found in the only US link in my short collection of sites was removed and replaced with a reference to a site that has the real thing. The out-of-date version was inadequate to support wusage.

Compatibility fixes for yet more compilers.

As always, if you have problems, contact me and I'll do my best to get you (and others with similar setups) up and running.

Upgrade Notes: ppmfig has not changed since version 2.0, but both wusage.c and usagegraph.c have changed in this version. So if you already have a working ppmfig you can forego rebuilding it. If you're having problems, though, be sure to try rebuilding ppmfig against an up-to-date pbmplus version.

#### What's New in version 2.1!

Version 2.1 now includes support for the CERN httpd as well as the NCSA

httpd. You should insert the line `NCSA_HTTPD` or `CERN_HTTPD` at the beginning of the `wusage.conf` file, as appropriate. (For backwards compatibility, `wusage` will assume NCSA if this line is absent.)

Version 2.1 corrects a bug in the use of wild cards: wild cards at the beginning of an entry in one of the exclusion lists now work properly (so entries such as `*.gif` are now correctly processed).

Version 2.1 now ignores white space at the end of entries in the exclusion lists; not strictly a bug in 2.0, but it saves a lot of grief.

What's New in version 2.0!

First, and most important- version 2.0 is now compatible with all, or nearly all, versions of Unix! Version 2.0 relied on certain time-handling routines that did not exist in non-Sun versions of Unix. These have been replaced.

Second, Version 2.0 supports the exclusion of unwanted accesses such as gif files, personal files, and other materials that distort the statistical picture of the server, in the opinion of the operator. This mechanism is entirely under the control of the operator-- no code changes are needed.

Third, a major bug resulting in incorrect top-ten lists when `wusage` attempted to take care of several unprocessed weeks in one pass was fixed. Credits and license terms `wusage 3.2` is copyright 1993, 1994, Quest Protein Database Center, Cold Spring Harbor Labs. Permission granted to copy and distribute this work provided that this notice remains intact. Modified versions should be cleared through Quest first; if this is not done, any modified version of the program must be clearly labeled as such.

The Quest Protein Database Center is funded under Grant P41-RR02188 by the National Institutes of Health.

Written by Thomas Boutell, 11/93 - 5/94.

The GIF code is based on that found in the `pbmplus` utilities, which in turn is based on `GIFENCOD` by David Rowley. See the notice below:

```
/*
** Based on GIFENCOD by David Rowley .A
** Lempel-Zim compression based on "compress".
**
** Modified by Marcel Wijkstra
**
** Copyright (C) 1989 by Jef Poskanzer.
**
** Permission to use, copy, modify, and distribute this software and its
```

\*\* documentation for any purpose and without fee is hereby granted, provided  
\*\* that the above copyright notice appear in all copies and that both that  
\*\* copyright notice and this permission notice appear in supporting  
\*\* documentation. This software is provided "as is" without express or  
\*\* implied warranty.  
\*\*  
\*\* The Graphics Interchange Format(c) is the Copyright property of  
\*\* CompuServe Incorporated. GIF(sm) is a Service Mark property of  
\*\* CompuServe Incorporated.  
\*/

What is wusage?

wusage maintains usage statistics for a WWW server. Specifically, it updates the following information, week by week:

- \* Total server usage, each week
- \* "Index" usage, each week (responses to ISINDEX pages)
- \* The top ten sites by frequency of access, each week
- \* The top ten documents accessed, each week
- \* A graph of server usage over many weeks
- \* An "icon" version of the graph for your home page!

What else do I need to use wusage?

To use wusage, you will need the following:

- \* The NCSA or CERN httpd World Wide Web server, or any "common logfile format" server
- \* A C compiler (any compiler should now work, even pre-ANSI K&R C)

That's it! previous versions required the presence of the pbmplus utilities and of a Unix shell. These requirements have been lifted. Version 3.2 should be a very easy (even trivial) port to MSDOS, including the GIF support routines. If you do this, please contact me so I can combine your code into the official package and make your binary available!

What if I don't use a standard server?

wusage is intended for use with the NCSA or CERN httpd servers, or with any server which produces the new "common logfile format". If you use a different server with a different access log file format, it will be necessary to patch the wusage.c source code appropriately, which should not be overly difficult. I will be glad to assist as best I can. Note that the author of your server should be using the new common log file format, so if they are not doing so I suggest you point this out to them.

How do I get wusage?

You can fetch wusage as a compressed tar file here. Or you can FTP it directly from [isis.cshl.org](http://isis.cshl.org), in the subdirectory `pub/wusage`.

How do I build wusage?

In order to build wusage, first untar the wusage.tar file with the following command:  
uncompress wusage3.2.tar.Z tar -xf wusage3.2.tar

This will create the directory "wusage3.2" beneath the current directory.

cd to this directory and examine the Makefile, which you may need to change slightly. Specifically, if you are using a different C compiler which is not named or aliased to "cc" (this is quite uncommon), change

CC=cc

to read

CC=acc

Or to another appropriate compiler.

If you are using the SGI C compiler, you will need to add "-cckr" to the CFLAGS line.

Now, to build the package, just type "make all". If all goes well, the program "wusage" will be compiled and linked without incident.

You have now built wusage. All that remains is to configure it for use with your server.

Configuring wusage for your server There are several parameters which must be set in order for wusage to properly interact with your server. These are set in the file wusage.conf. A sample wusage.conf file is included in the tarfile, and you can use this file as a starting point. You will definitely need to edit this file to configure wusage properly for your server unless it is identical to ours.

Here is the sample wusage.conf file. Note that lines beginning with "#" are comments and are ignored. Note also that blank lines are NOT considered comments and should be avoided.

Important change in version 3.2: the "home page" line has been removed. Delete it from your wusage.conf and add the prefix and suffix lines (see below).

```
#Type of server log: COMMON (all new servers), NCSA_HTTPD, CERN_HTTPD,  
#or PLEXUS_HTTPD.  
#The latter three are for older versions of those servers; newer versions  
# should use the COMMON log file format (but CHECK YOUR DOCUMENTATION).  
COMMON  
#Name of your server as it should be presented  
Quest  
#File to use as a prefix; MUST BE A COMPLETE FILE SYSTEM PATH. REALLY.  
#NOT A URL.  
/home/www/prefix  
#File to use as a suffix; MUST BE A COMPLETE FILE SYSTEM PATH. REALLY.
```

```

#NOT A URL.
/home/www/suffix
#Directory where html pages generated by usage program should be located
/home/www/web/usage
#URL to which locations of html pages should be appended for usage reports
#(the same as the first line, but in web space, not filesystem space)
/usage
#Path of ncsa httpd log file
/home/www/ncsa/logs/access_log
#Your top-level domain name (org, edu, com... just the topmost level)
org
#Hidden items
#Ignored items
#Ignored sites
#Domain aliases or "none"
#none
    {
        aliasname
        domain
        domain
        domain
    }
    ... More aliases, if any ...

```

The first non-comment line should read:

COMMON or NCSA\_HTTPD or CERN\_HTTPD or PLEXUS\_HTTPD

as appropriate to your server's log file format. Note that the latest versions of CERN and NCSA servers produce the COMMON log file format, and setting this line to a different value won't work for those versions!

#### UPPERCASE REQUIRED.

Note to those upgrading: once you switch to the COMMON log file setting, wusage can't read any data in the old format that may be lying around, but it can skip over it tactfully. The upshot of this is that if you've been running wusage all along, you'll simply be able to start using it again and will only need to adjust the results for the one week during which you made the changeover to a common-logfile-format server version.

For those using wusage for the first time, this is a thornier problem, but it can be handled with some ingenuity (by switching the setting of the first line after running wusage on the pre-common format part of your log, then deleting the older content). I encourage server authors (and anyone else for that matter!) to write a conversion filter to translate old-style log file formats to the new style. It shouldn't be very difficult. At worst, you'll have statistics only from the point at which you switched to a common-logfile-format server.

The second non-comment line should contain the name of your server as you would like it to be referred to in the usage page.

The third line should contain the full filesystem path (NOT URL) of a file you would like to have copied in at the beginning of each page generated by wusage, or the word none (in

lowercase letters). You can use this mechanism to add a link up to your home page, or an illustration of your choice.

The fourth line is just like the third, but specifies a suffix file to be appended at the end of each page.

Sample prefix and suffix files are provided. Note the link to the wusage documentation in the suffix file. You are not required to keep this link, but we will greatly appreciate it if you do so. (Of course, if your site is strictly internal and behind a firewall, you should remove the link, since it won't work for your users.)

The fifth line contains the directory in your file system in which html pages generated by wusage should reside. This will usually be a subdirectory of your server root directory called "usage". (In our case, `DOCUMENT_ROOT` is `/home/www/web`, so the fifth line is `/home/www/web/usage`.)

**IMPORTANT:** this directory should not be shared with other information! Please give usage a subdirectory to itself, since it creates and deletes files fairly freely and assumes its directory is a safe place in which to do so.

The sixth line is the "base URL" for html pages generated by wusage. This is similar to the second line, but is the location in web space, not in filesystem space. Thus, if `DOCUMENT_ROOT` is `/home/www/web` and you set the second line to `/home/www/web/usage`, the fourth line should be set to just `/usage`.

The seventh line is the location of the NCSA server access\_log file, which wusage needs to be able to read in order to compute statistics. This file is located in `.../ncsa/logs`; `...` is the location at which you installed the server. In our case it is installed beneath `/home/www`.

The eighth line is the default domain, which should be the domain in which your own server is located. For instance, if your server's name is `siva.cshl.org`, this line should read `org`. This is new in version 3.1 and later.

#### Excluding unwanted accesses

The lines above are followed by four lists of items, enclosed by { and } characters. By default, these lists are empty. The absence of the lists is tolerated for backwards compatibility.

The first is a list of items which should be "hidden". This means that they will still register in the total number of accesses, but they will never be in the top ten for any week.

The second is a list of items which should be "ignored". These items never appear in the total number of accesses OR in the top ten-- they are completely ignored.

The third is a list of sites to be ignored. This is useful if many of the accesses to your server are made by you personally and you are more interested in counting accesses made by other sites.

For instance, if you want to keep .gif files (frequently inline) out of the top ten, completely ignore files coming from users' personal directories, and ignore accesses from your own site "here.com", the three lists would look as follows. (Note that asterisks are acceptable as wild

cards, just as they are in the file system; question marks are also acceptable to substitute for any single character.)

```
#Hidden items
*.gif
#Ignored items
/~*
#Ignored sites
www.ourcompany.com
*.ourschool.edu
```

This mechanism makes it much easier to arrive at a meaningful top-ten list.

#### Charting access by domain

Pie charts showing the usage of your server by domain, telling you where in the world people are connecting to your server from, are now available. These pie charts appear on each week's page.

To make them more useful, it is possible to combine countries into continent domains.

The last section of the `wusage.conf` file is now made up of continent aliases. Alternatively, you can turn off domain charts altogether by uncommenting the "none" line just before the continent aliases.

The continent aliases provided work well, but if you would like to alter them (to add new countries or break up continents, for instance, if your server is located in Europe), here are the rules:

The entire set of aliases is enclosed in a `{ ... }` pair.

Each alias is enclosed in a `{ ... }` pair (see the example set in `wusage.conf`). The first domain in each alias is the name that the rest will be aliased to. This adds them together to make the result show up better in the pie chart and the list of the top ten domains. The first domain can itself be a real domain (such as the little-used "us" domain, to which you could additionally alias gov, edu, org, mil and com, though this is not always correct), or it can be a made-up domain such as "Asia". Domain names are generally kept short so they will fit into the pie charts well.

See the provided `wusage.conf` file for examples.

The pie chart only shows domains which take up a sufficient percentage to be legible in the chart, but the top ten list always shows the top ten domains (if there have been accesses from ten or more domains).

The "?" domain is assigned to accesses from sites whose names are unknown.

The default domain (line seven of `wusage.conf`) is assigned to sites which have no periods in their names (ie, they are assumed to be local sites in your own domain). The "other" category in the pie chart is assigned to all accesses from domains too small to show up in the chart.

Again, if the pie charts don't work well for you (because all of your accesses are from one domain, or because your nameserver is broken and you only have IP addresses in your access logs), you can shut them off by uncommenting "none" in your wusage.conf file.

## Running wusage

There are three common ways to run wusage: as an automatic weekly job, by hand, and through a cgi script (which allows you to have a "button" on one of your web pages to update the information).

An automatic weekly job is the best approach, since this is the frequency with which wusage generates reports. If you are using a Unix system, it is easy to do this using the program "cron".

wusage needs to be run on a weekly basis in order to keep useful statistics. Specifically, it should be run as soon after midnight on Sunday as possible. For the purposes of creating an html report, wusage should always be run with this ONE option (this is a change from versions before 3.0):

-c (location of wusage.conf file) which specifies the location of the configuration file.

You can simply run wusage by hand with the -c option (example: wusage -c wusage.conf). You will need to do so once a week.

To run it from a cgi script, create a cgi script which executes the above command and echoes back a reasonable page to the user indicating success. (Since reports are weekly no matter how often the program is run, it is recommended that such a button be placed on a private page, since it has no dramatic effect and need not be run incessantly by users.)

In order to install wusage as a regularly-scheduled automatically-run program, you need to add it to your crontab file and submit it to the program "crontab". Our crontab file looks like this:

```
1 0 * * 0 /home/www/wusage -c /home/www/wusage.conf
... other jobs, if any ...
```

The crontab file submitted to the Unix system with the following command, assuming it is called "crontab.txt": crontab crontab.txt

Of course, if you run the www server as root, you no doubt already have a crontab file for root, to which you will want to add this line, following this with a reinstall using crontab. (We created a separate www account to facilitate this sort of thing; I recommend this strategy to other server administrators.)

## Hooking up wusage

Everything else is taken care of; all that remains is to run wusage for the first time (to make sure the various html and .gif files actually exist) and linking the usage report to your home page. Run wusage by hand using the following command:

```
wusage -c /home/www/wusage.conf
```



(Substitute the directory where wusage.conf resides on your system for /home/www in the above.)

Now, if all has gone well, edit your home page to include a link to the usage report. Here is the relevant excerpt from our home page:

```
<P>Usage of the Quest WWW server is kept track of through <A  
HREF="/usage/index.html"> <IMG ALIGN=TOP SRC="/usage/usage.graph.small.gif"></A>  
<A HREF="/usage/index.html">usage statistics</A>.</P>
```

In addition to obvious name changes, you may need to change the directory linked to if you did not use /usage in your configuration file.

Note that in addition to a normal text link, a small usage graph is provided as an icon. This graph is genuine-- it is updated at the same time as the larger graph on the main usage page!

#### Purging access\_log (how and why)

Your access\_log file will grow tremendously over time, particularly if your server is heavily used. It is desirable to purge this file periodically, and this can be done provided you follow these directions.

Take note of the most recent week for which wusage has generated a complete report. Determine the date on which this week ended (the usage report displays the date the week began).

Now edit your access\_log file and find the first entry that falls AFTER the completion of that week. It is safe to delete all entries BEFORE that line in the access\_log file.

Important note:if you do purge your access\_log file, then be sure to back up the directory in which wusage keeps its html pages. This directory contains important summary information for previous weeks which wusage must have in order to graph information regarding past weeks no longer in the access\_log file. Of course, you should also compress and back up your old access\_log data.

Please tell us you're running wusage!

When you contact us and let us know you are running wusage, you help us justify the time spent in maintaining and improving it. So please let us know. If you can provide a URL for your usage page, that's great, but if it's an internal server (not open to the public), a simple note is just as welcome.

If you have problems If you have any difficulties with wusage, feel free to contact the author, Thomas Boutell.

---

webmaster@chem.leeds.ac.uk

## 2.2.- GETSTATS Version 1.2

What is getstats?

Getstats (formerly called getsites) is a versatile World-Wide Web server log analyzer. It takes the log file from your CERN, NCSA, Plexus, GN, MacHTTP, or UNIX Gopher server and spits back all sorts of statistics. It was originally written according to suggestions as given by Tony Sanders (sanders@bsdi.com).

Getstats is not in the public domain. However, you may distribute it or its supporting files and documents in any format freely as long as it is unchanged from the original distribution.

\* In order to use getstats, you must agree to the terms of this license.

-- Kevin Hughes, kevinh@eit.com

1. Major report types
2. Usage
3. Command-Line Options
4. Internal Options
5. The Getstats Form Interface
6. Frequently Asked Questions
7. Distribution and Compilation Notes
8. Other Statistics Gathering Programs

---

### Major Report Types

Currently there are twelve major types of reports this program can produce. You can use as many options as you like to create combinations of reports.

1. getstats -c (concise report)

#### HTTP Server General Statistics

Server: <http://www.eit.com/> (NCSA)

Local date: Fri Feb 11 18:17:07 PM PST 1994

Covers: 02/09/94 to 02/11/94 (3 days).

All dates are in local time.

Requests last 7 days: 4495

New unique hosts last 7 days: 358

Total unique hosts: 358

Number of HTML requests: 1854

Number of script requests: 472

Number of non-HTML requests: 2169

Number of malformed requests (all dates): 5

Total number of all requests/errors: 4500

Average requests/hour: 90.2, requests/day: 2164.7

Running time: 11 seconds.

This basic set of statistics is always output when getstats runs. Using the -c option will only produce this statistics paragraph.

## 2. getstats -m (monthly report)

### HTTP Server Monthly Statistics

Covers: 10/30/93 to 11/08/93 (9 days).

All dates are in local time.

Each mark (#) represents 1000 requests.

-----  
Oct (10/30/93): 569 : #

Nov (11/04/93): 2 :

...

The -m option will produce a monthly report of server use. The dates in the report are the first day of reported activity for that month.

## 3. getstats -w (weekly report)

### HTTP Server Weekly Statistics

Covers: 12/28/93 to 01/27/94 (32 days).

All dates are in local time.

Each mark (#) represents 500 requests.

-----  
Week of 12/27/93: 1878 : ###

Week of 01/03/94: 5606 : #####

Week of 01/10/94: 23287 : #####

...

The -w option will produce a weekly report of server use. The dates in the report are always the Monday of that particular week.

## 4. getstats -ds (daily summary)

### HTTP Server Daily Summary

Covers: 12/28/93 to 01/27/94 (32 days).

All dates are in local time.

Each mark (#) represents 1000 requests.

-----  
Mon: 16018 : #####

Tue: 13219 : #####

Wed: 9904 : #####

...

The -ds option produces a daily summary, which shows the aggregate number of requests for a particular day of the week.

## 5. getstats -d (daily report)

### HTTP Server Daily Statistics

Covers: 12/28/93 to 01/27/94 (32 days).

All dates are in local time.

Each mark (#) represents 100 requests.

```
-----  
12/28/93 (Tue): 88 :  
12/29/93 (Wed): 258 : ##  
12/30/93 (Thu): 591 : #####  
12/31/93 (Fri): 775 : #####  
...
```

The -d option produces a daily report, which shows the number of requests per day and the date.

## 6. getstats -hs (hourly summary)

### HTTP Server Hourly Summary

Covers: 12/28/93 to 01/27/94 (32 days).

All dates are in local time.

Each mark (#) represents 200 requests.

```
-----  
midnite: 1266 : #####  
1:00 am: 1206 : #####  
2:00 am: 1238 : #####  
...
```

The -hs option produces an hourly summary, which shows the aggregate number of requests for a particular hour.

## 7. getstats -h (hourly report)

### HTTP Server Hourly Statistics

Covers: 12/28/93 to 01/27/94 (32 days).

All dates are in local time.

Each mark (#) represents 20 requests.

```
-----  
12/28/93 (Tue)  
  
3:00 pm: 39 : #  
4:00 pm: 12 :  
5:00 pm: 36 : #  
...
```

The -h option produces an hourly report, which shows the number of requests per hour, the day of the week, and the total number of requests for each day.

#### 8. getstats -f (full report)

##### HTTP Server Full Statistics

Sorted by number of requests.

Covers: 12/28/93 to 01/27/94 (32 days).

All dates are in local time.

# of Requests : Last Access (M/D/Y) : Hostname

```
-----  
6994 : 01/26/94 : kmac  
1751 : 01/26/94 : eitech  
1096 : 01/27/94 : jhvh-1  
...
```

The -f option tells getstats to create a full report sorted by host name (and IP address). Use the -fa option to make a full report sorted by the number of accesses, the -fd option to create a full report sorted by the last access date, or the -fb option to create a full report sorted by the number of bytes transferred.

#### 9. getstats -r (request report)

##### HTTP Server Request Statistics

Sorted by number of requests, 1560 unique requests.

Covers: 12/28/93 to 01/27/94 (32 days).

All dates are in local time.

# of requests : Last Access (M/D/Y) : Request

```
-----  
4260 : 01/27/94 : /eit.home.html  
3330 : 01/27/94 : /graphics/stripe.bottom.gif  
2831 : 01/27/94 : /graphics/ball.black.gif  
...
```

The -r option tells getstats to create a report of requests sorted by the request name. Use the -ra option to sort by accesses, -rd to sort by the last access time, -rb to sort by the number of bytes transferred, and -rf to sort by individual file sizes.

#### 10. getstats -dn (domain report)

##### HTTP Server Domain Statistics

1 level, sorted by domain name, 22 unique domains.

Covers: 02/09/94 to 02/10/94 (2 days).

All dates are in local time.

# reqs : # uniq : Last Access (M/D/Y) : Domain

```
-----  
180 : 28 : 02/10/94 : (numerical domains)  
27 : 1 : 02/10/94 : .at  
28 : 3 : 02/10/94 : .au
```

22 : 2 : 02/10/94 : .ca

The -dn option generates a domain report, sorted by domain name. Use -da to sort by the number of requests, -dd to sort by last access date, -db to sort by the number of bytes transferred, or -du to sort by the number of unique domains. The unique domain number is the total number of unique sites under a domain. In the example above, for instance, a total of 3 unique sites came from the .au domain.

#### 11. getstats -dt (directory tree report)

##### HTTP Server Tree Report

Covers: 12/28/93 to 01/07/94 (12 days).

All dates are in local time.

# of Requests : Last Access (M/D/Y) : Dir/File

---

```
55 : 01/07/94 : /reports
51 : 01/07/94 : /ht93
562 : 01/07/94 : /demos
487 : 01/07/94 : /asiceda
```

The -dt option generates a directory tree report, which cannot be sorted. The number of requests and last request date for directories and files is displayed. The request count for directories is the amount of requests for that directory plus the sum of all requests for the files and subdirectories under it.

If you find this report is empty, try using the -dr option without specifying a directory. This will tell getstats to make a tree report without verifying that the files and directories reported in the log file actually exist.

#### 12. getstats -e (file) (error report)

##### HTTP Server Error Report (All Dates)

---

```
kmac [Thu Dec 30 23:20:21 1993] get / foo
kmac [Thu Dec 30 23:20:37 1993] get foo /
kmac [Thu Dec 30 23:20:55 1993] get http://www.eit.com/ foo
```

-e generates a report of all malformed (or ignored) requests for all dates in the order they were encountered in the log file. If a filename is given as the argument to the option, bad requests will be appended to an error file, where they can be analyzed later.

#### getstats -a (all reports)

The -a option will produce all of the above reports, with list reports sorted by the number of accesses, if possible. If you want a report sorted another way, however, specify the correct option after the -a flag.

example: getstats -a -fb

This will create all reports sorted by number of requests, with the exception of the full report, which is sorted by byte traffic, and the error report, which must be specified on the command line.

---

#### Usage

usage: getstats [-C,-N,-P,-G,-A,-O], -M, -c, -m, -w, -ds -d, -hs, -h,  
-e ["file"], -a, -dt, [-f,-fa,-fd,-fb], [-r,-ra,-rd,-rb,-rf],  
[-dn,-da,-dd,-db,-du], -dl #, -df "file",  
-sa "string", -ss "string", -sr "string", -sp "string",  
-sd "string", -sh "string", -sw "string",  
-b, -i, -ip, -p, -ht, -t #, -dr ["dir"], -l "file"

options: No option gives the default report.

- C, -N, -P, -G, -A, -O
  - : use CERN, NCSA, Plexus, GN, MacHTTP,  
or UNIX Gopher server log format
- M : use common logfile format
- c, -m, -w, -ds, -d, -hs, -h, -e, -a
  - : concise, monthly, weekly, daily summary, daily, hourly summary,  
hourly, error, and all reports
- f, -fa, -fd, -fb : full report
  - : sorted by address, accesses, date, or bytes
- r, -ra, -rd, -rb, -rf : file request report
  - : sorted by request, accesses, date, bytes, or file size
- dn, -da, -dd, -db, -du : domain report
  - : sorted by domain, accesses, date, bytes, or unique domains
- dl : number of domain levels to report
- df : file to look up domain codes from
- dt : directory tree report
- sa, -ss, -sr, -sp : filter log by "string"
  - : only addresses, skip addresses, only reqs, skip reqs
- sd : report entries with date "m/d/y"
- sh : report entries with hour "h"
- sw : report entries with day "day"
- b : add byte traffic statistics to all reports
- i : take input from standard input
- ip : look up all IP addresses
- p : display progress meter
- ht : produce HTML output
- t : take top # lines of list reports
- dr : root Web/Gopher directory
- l : logfile to use

docs: <http://www.eit.com/software/getstats/getstats.html>

To see the usage, run getstats with a -z option.

---

Command-Line Options -dr directory, -l file, -C, -N, -P, -G, -A, -O, -M (root directory, logfile

and log type). The `-dr` option tells `getstats` what your root Web or Gopher directory is. This information is needed in order to determine byte statistics.

example: `getstats -dr "/usr/local/www"`

The `-l` option specifies the log file to use. The `-C`, `-N`, `-P`, `-G`, `-A`, and `-O` options will tell `getstats` to expect the log file to be in either CERN, NCSA, Plexus, GN, MacHTTP, or UNIX Gopher format.

example: `getstats -l my.ncsa.log -N`

example: `getstats -l my.plexus.log -P`

The `-M` option tells `getstats` to expect the log to be in the "common" log file format, a standard that was agreed upon by the World-Wide Web community. If your log looks something like this:

`www.eit.com - - [01/Jan/1994:10:30:00 +0000] "GET /test.html" 200 123` then your server is using the format. Include the option in the command line:

example: `getstats -l cern.common.log -C -M`

example: `getstats -l ncsa.common.log -N -M`

If you use NCSA's `httpd` 1.2 or later, or CERN's server version 1.16b or later, you are probably using the common log format.

`-sa` string, `-ss` string, `-sr` string, `-sp` string (address and request masks)

The `-sa` option will only report (IP or name) addresses matching the conditions in the mask string. The `-ss` option will skip addresses matching the string conditions. `-sr` will only report requests matching the string conditions, and `-sp` will skip report requests matching the string conditions.

For these four case-insensitive string masks, the following rules apply:

1. You can use asterisks as wildcards in specifying the string, at each or both ends of the string.
2. Masks without wildcards must match exactly.
3. You can make lists of masks, separated by commas.

example 1: `getstats -sa "*.com, *.edu" -ss "*.eit.com"`

example 2: `getstats -sr "*.html, *.gif" -sp "**secret**"`

example 3: `getstats -sa ".*" -sp "/internal/demo.html"`

example 4: `getstats -ss "dopey, sneezy, grumpy"`

1. Sites from ".com" and ".edu" domains are reported, but anything coming from "eit.com" is skipped.
2. HTML and GIF requests are reported, but any requests with the word "secret" in them are skipped.
3. Addresses with a period in them are reported (this is useful for filtering local machines), and any requests exactly matching "/internal/demo.html" are skipped.



4. Requests coming from "dopey", "sneezy", and "grumpy" are skipped. -sd string, -sh string, -sw string (date, hour, and day masks) The -sd option reports requests matching the date conditions in the string.

In a similar way, the -sh option filters by hour, and the -sw option filters by the day of the week.

For these three case-insensitive string masks, the following rules apply:

1. The date string must be in the format "m/d/y". Ranges in brackets, such as "[xx-xx]" can be applied, and asterisks as wildcards can be used only in place of an entire field.

2. Other valid specifications for the date string: today, yesterday, thisweek, lastweek, thismonth, and lastmonth.

3. The hour string must contain numbers from 0 to 23, with 0 being the hour from midnight to 1 am, and 23 being 11 pm. Ranges such as "xx-xx" can be specified.

4. The day string must be a three-letter day of the week, such as "mon", "tue", and "wed". Ranges such as "xxx-xxx" can be specified.

5. Other valid specifications for the day string: weekends and weekdays.

example 1: getstats -sd "\*/[4-10]/93" -sh "6" -sw "mon"

example 2: getstats -sd "1/[5-30]/1993" -sh "5-17" -sw "wed-sun"

example 3: getstats -sd "[1-5]\*/[91-94]" -sh "-17" -sw "-thu"

example 4: getstats -sd "lastweek" -sh "15-" -sw "tue-"

1. Only accesses that occurred from the 4th to the 10th on any month in 1993 on a Monday at 6 pm will be reported.

2. Requests from January 5th to the 30th, 1993 are reported, as are requests from 5 am to 2 pm, on Wednesdays to Sundays.

3. Requests from January to May 1991 to 1994 are reported, as are requests from midnight to 2 pm, on Mondays to Thursdays.

4. Requests from last week are reported, as are requests from 3 pm to 11 pm, on Tuesdays to Sundays.

-i, -p, -ht (input and output)

The -i option will allow you to take input from standard input, so you can do things such as piping lines into getstats. This option is disabled for VMS platforms.

example: tail -100 my.log | getstats -

-p displays a progress meter, so you can see where getstats is in its processing. The -ht option will report all reports in a single-page HTML format, with appropriate links to server support URLs and this documentation.

-dl number, -df file (domain report options)

The -dl option allows you to specify how many domain levels to report. For instance, with the number of levels set at 1, the domain report would look something like:

27 : 1 : 02/10/94 : .at

28 : 3 : 02/10/94 : .au

22 : 2 : 02/10/94 : .ca

18 : 2 : 02/10/94 : .ch

640 : 37 : 02/10/94 : .com

With the number of levels set at 2:

27 : 1 : 02/10/94 : .at  
27 : 1 : 02/10/94 : .at.ac  
28 : 3 : 02/10/94 : .au  
28 : 3 : 02/10/94 : .au.edu  
22 : 2 : 02/10/94 : .ca  
18 : 1 : 02/10/94 : .ca.nrc  
4 : 1 : 02/10/94 : .ca.uwo

The -df option allows you to specify a file with descriptions for domain codes, to make the domain report a bit easier to understand. A file with descriptions is available by FTP at [ftp.eit.com](ftp://ftp.eit.com/pub/web/software/getstats), in the /pub/web/software/getstats directory. You can also get the list by selecting this.

With domain descriptions:

27 : 1 : 02/10/94 : Austria (.at)  
28 : 3 : 02/10/94 : Australia (.au)  
22 : 2 : 02/10/94 : Canada (.ca)  
18 : 2 : 02/10/94 : Switzerland (.ch)  
645 : 37 : 02/10/94 : US Commercial (.com)  
13 : 2 : 02/10/94 : Germany (.de)

-b, -ip, -t number (byte reporting, IP lookup, and top lines)

The -b option will report byte traffic statistics in all reports, and an extra column in list reports will be added for byte traffic. In addition, the average number of bytes transferred per hour and day will be added to the statistics paragraph. Byte counts for each file are determined by getting the size of the requested file, the path of which is determined by the top web directory and the request. However, like many log analyzers, there are byte statistics that getstats can't report:

- \* Redirected URLs
- \* The sizes of files in personal HTML directories
- \* The sizes of scripts
- \* Old files that have been changed
- \* Requests that have been rejected (perhaps due to access control)

Counts for these requests, or counts for requests that can't be determined, are reported as zero.

The -ip option will make getstats attempt to look up host names from IP addresses. This feature slows processing time, but is useful in analyzing logs from servers that don't look up IP addresses, such as the CERN server.

The -t option allows you to specify how many top lines to report in full, request, and domain reports. Using this, you can easily generate "Top 10" lists and short summaries.

---

## Frequently Asked Questions

Getstats crashes and burns when I run it.

First, double-check that you've specified the correct format for your log file. If it's in the "common" format, add the -M option to the command line. If all else fails, and you think there may be a bad line in your log file, try to narrow down the section under which getstats crashes and email it to me so I can test it out.

How can I use getstats on multiple log files?

A good idea is to use the -i option on previously archived logs, for instance:

```
zcat log.*.Z | getstats -i
```

Will getstats be rewritten in perl?

No.

How do I run getstats as a CGI script?

First, uncomment the "#define CGI" line in the source code and compile. To call getstats from a URL with options, use a question mark after the program name, plus characters for spaces, and "%22" for quotes in the command line, such as:

Normal example: `getstats -a -fb`

URL example: `http://www.eit.com/cgi-bin/getstats?-a+-fb`

Normal example: `getstats -ss "*.com"`

URL example: `http://www.eit.com/cgi-bin/getstats?-ss+%22*.com%22`

Why is my directory tree report empty?

If an actual root Web directory is specified, nothing will be reported if getstats can't find the files. Use the -dr option with no arguments to generate a report if no physical directory exists.

Why are some files reported as directories in my tree report?

Unless you tell getstats where the actual files exist, it can only make a "best guess" as to whether a file is a directory or not. Use the -dr option with your root Web directory to tell it where your files are. See the above documentation for more details.

The reported errors from my "common" format log file look OK to me!

The second to last field on each line indicates the status code the server returned. Codes not in the "2xx" range mean some sort of error occurred. You can get a list of these codes at <http://info.cern.ch/hypertext/WWW/Protocols/HTTP/HTRESP.html>.

Can I run getstats with my Windows NT server?

It appears that the EMWACS Windows NT server uses the CERN format. For their server, include the -C option in the command line.

---

## Distribution and Compilation Notes

---

## Internal Options

Every option available from the command line can be hardcoded into getstats, so gigantic series of options don't have to be used all the time!

Getstats has a number of options that are not available from the command line which must be specified in the source code:

- \* The root directory of your web or gopher tree, the URL for your server, and aliases for empty and slash ("/") requests.
- \* The mark character for graphs, and the number of requests and bytes per mark for different reports. The default settings have been tuned for light server use, or about 200 to 1,000 requests per day.
- \* The character length to truncate graph reports and requests reports to.
- \* The option to display all dates in local or GMT time (the default is GMT time), and what time zone the logfile uses.
- \* Whether or not you want to show numerical domains in the domain report.
- \* Whether or not you want to report hours, minutes, and seconds of the last access time in non-graph reports.
- \* Whether or not you wish to show files and their related statistics as well as directories in directory tree report.
- \* Whether or not to check the actual files to see if they exist in the directory tree report.
- \* For HTML request reports, whether or not to display requests as selectable URLs.
- \* If you have a GN server, whether you want to report both Gopher and HTML requests or not.
- \* Whether or not you want to display an image in your HTML reports (like the one on this page).
- \* How dates are to be displayed - either as M/D/Y or D/M/Y.

---

## The Getstats Form Interface

Thanks to Brian Behlendorf (brian@gw.wired.com), there's now a pretty nifty HTML form interface to getstats. With it, you can choose to have the results mailed back to you or displayed within your Web browser.

You can get the C program statform.c and the HTML form statform.html by FTP at ftp.eit.com in the /pub/web.software/getstats directory. Or download statform.c and statform.html directly.

To start using the form:

1. Compile getstats with the "CGI" option (you'll need to edit the source code for this).
2. Edit statform.c, specifying the default log file to analyze, the root directory of your Web or Gopher tree, and the path to your getstats program.
3. Compile statform.c and put the executable where you store your other CGI programs.
4. Edit statform.html and specify the URL that will execute the statform program.
5. Make sure both statform and getstats are executable by your server.
6. Load up statform.html and have fun!

This program compiles fine under gcc and runs under SunOS 4.1.3 (Solaris 1.1) with no problems. Of course, you'll have to define the internal options before you compile it.

You can get the program by FTP'ing ftp.eit.com and going to the /pub/web.software/getstats directory. Some files you'll need for compiling under VMS, as well as the getstats icon (which you can use in your HTML reports) are there also.

Other environments that getstats has reportedly compiled and run under with no problems are Solaris 2.3, DEC Ultrix, DEC OSF/1 V1.3 in Alpha AXP, SGI with Irix 4.0, AIX, NeXTstep, HPUX, A/UX, and VMS (with or without UCX). If you have compiled successfully on another platform or experience problems, please let me know.

Wanted!

- \* Suggestions for features that are most needed for administrators.
- \* Ways I can speed up getstats and/or make it more efficient.
- \* Comments, improvements, patches, and suggestions are greatly appreciated - you can send them all to kevinh@eit.com.

---

#### Other Statistics Gathering Programs

Of course, getstats doesn't do everything, and depending on your needs you may want to look into the following programs:

- \* A package that generates graphs from getstats output.
- \* WWWstat, a full-featured NCSA server log analyzer written in Perl.
- \* wusage, a C program that can generate graphs and nifty graph icons along with basic CERN and NCSA log statistics.
- \* WebReport, a Perl NCSA log analyzer that comes with NCSA's httpd.
- \* gnlog, a Perl GN log analyzing program.
- \* glog, a Gopher log analyzer that can sort statistics in many ways.

This and other Gopher utility software is available at the boombox.micro.umn.edu Gopher.

## 2.3.-WWWSTATS

wwwstat: Process a sequence of NCSA httpd 1.2 access\_log files and output a summary of the access statistics in a nice HTML format. The program oldwwwstat handles NCSA httpd 1.1 and earlier.

Copyright (c) 1994 Regents of the University of California.

---

### Licensing and Distribution Information:

This software has been developed by Roy Fielding <fielding@ics.uci.edu> as part of the Arcadia project at the University of California, Irvine. Wwwwstat was originally based on a multi-server statistics program called fwgstat-0.035 by Jonathan Magid (jem@sunsite.unc.edu) which, in turn, was heavily based on xferstats (packaged with the version 17 of the Wuarchive FTP daemon) by Chris Myers (chris@wugate.wustl.edu).

Those parts of wwwstat derived from fwgstat and xferstats are in the public domain. As such, this software and all derivations will always be free to the general public.

The latest version of wwwstat can always be obtained at

<http://www.ics.uci.edu/WebSoft/wwwstat/>  
or by anonymous ftp from  
<ftp://liege.ics.uci.edu/pub/arcadia/wwwstat/>

The wwwstat package and those portions developed exclusively at the University of California are covered by the above copyright notice.

Redistribution and use in source and binary forms are permitted, subject to the restriction noted below, provided that the above copyright notice and this paragraph and the following paragraphs are duplicated in all such forms and that any documentation, advertising materials, and other materials related to such distribution and use acknowledge that the software was developed in part by the University of California, Irvine. The name of the University may not be used to endorse or promote products derived from this software without specific prior written permission. THIS SOFTWARE IS PROVIDED "AS IS" AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, WITHOUT LIMITATION, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE.

Use of this software in any way or in any form, source or binary, is not allowed in any country which prohibits disclaimers of any implied warranties of merchantability or fitness for a particular purpose or any disclaimers of a similar nature.

IN NO EVENT SHALL THE UNIVERSITY OF CALIFORNIA BE LIABLE TO ANY PARTY FOR DIRECT, INDIRECT, SPECIAL, INCIDENTAL, OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OF THIS SOFTWARE AND ITS DOCUMENTATION (INCLUDING, BUT NOT LIMITED TO, LOST PROFITS) EVEN IF THE UNIVERSITY OF CALIFORNIA HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

---

Installation instructions:

1. Get the wwwstat package from the distribution site (above). Normally, it will be in the form of a compressed unix tar file. If it has not already been decompressed by your WWW client, then do one of:

```
% uncompress wwwstat-1.0.tar.Z
% gunzip wwwstat-1.0.tar.gz
```

depending on which compressed version you downloaded.

2. Move the resulting wwwstat-1.0.tar file to the directory above where you want to install wwwstat, cd to that directory, and do % tar xvf wwwstat-1.0.tar to create the directory ./wwwstat-1.0 containing the following:

```
Changes      -- the list of known problems and version information
README       -- this file
country-codes -- a table of Internet domains and their country names
example.html  -- an example of what wwwstat output should look like
old2newlog    -- A tool for converting httpd 1.1 logs to 1.2 format
oldwwwstat    -- wwwstat for old NCSA httpd 1.0 or 1.1 servers
wwwstat       -- the main perl script
```

If you are already using NCSA httpd 1.2, delete the oldwwwstat script.

3. Configure the wwwstat script(s) to match the server configuration and default options desired for your site. You will probably need to change the following (with any text editor).

The first line (it should point to your perl executable) #!/usr/public/bin/perl

The following variables set in the first section of code:

```
$OutputTitle    # The output document's HTML Title.
$LastSummary    # The URL of the previous summary period
$ServerHome     # The server's default home page.
$countrycodefile # The location of the country-codes file.
$access_log     # The location of your default server access log
$srml_conf      # The location of your server configuration file
$zcat           # The name of your "uncompress to stdout" program
$zhandle        # The file extensions that indicate compressed
$AppendToLocalhost # If address in log entry is one word (a local host),
                  # what should be appended? (e.g. ".sub.dom.ain")
$mydom1         # Identify the last two components of your local
$mydom2         # hostname addresses for special treatment
$HeadEstimate   # Estimated size of the response headers
```

The defaults for options that can be overridden by the command line:

```
$LocalFullAddr  # Show full address for hosts in my domain?
$OthersFullAddr # Show full address for hosts outside my domain?
$ShowUnresolved # Show all unresolved addresses?
```

4. Make sure the script is executable:

```
% chmod 755 wwwstat
```

5. That's it. You should now be able to run wwwstat, e.g.

```
% wwwstat > results.html
```

6. If you have some old (prior to 1.2) logfiles that you want converted to the new format, you will also need to customize the old2newlog script (most variables are the same as those above). Usage information can be obtained via the -h option.

---

---

Usage: (NOTE - oldwwwstat has a different set of options)

```
wwwstat [-helLoOuUrvx] [-s srmfile] [-i pathname]
        [-a IP_address] [-c code] [-d date] [-t hour] [-n archive_name]
        [-A IP_address] [-C code] [-D date] [-T hour] [-N archive_name]
        [logfile ...] [logfile.gz ...] [logfile.Z ...]
```

Display Options:

- h Help -- just display this message and quit.
- e Display all invalid log entries on STDERR.
- l Do display full IP address of clients in my domain.
- L Don't (i.e. strip the machine name from local addresses).
- o Do display full IP address of clients from other domains.
- O Don't (i.e. strip the machine name from non-local addresses).
- u Do display IP address from unresolved domain names.
- U Don't (i.e. group all "unresolved" addresses under that name).
- r Display table of requests by each remote ident or authuser.
- v Verbose display (to STDERR) of each log entry processed.
- x Display all requests of nonexistent files to STDERR.

Input Options:

- s Get the server directives from the following srm.conf file.
- i Include the following file (assumed to be a prior wwwstat output).
- ... Process the sequence of logfiles (compressed if extension (gz|Z|z)).

Search Options (include in summary only those log entries):

- a Containing a hostname/IP address matching the given perl regexp.
- A Not containing " " " " " " " " " "
- c Containing a server response code matching the given perl regexp.
- C Not containing " " " " " " " " " "
- d Containing a date ("Feb 2 1994") matching the given perl regexp.
- D Not containing " " " " " " " " " "
- t Containing an hour ("00" -- "23") matching the given perl regexp.
- T Not containing " " " " " " " " " "
- n Containing an archive (URL) name matching perl regexp (except +.).
- N Not containing " " " " " " " " " "

Note that the Search Options allow for full use of perl regular expressions (with the exception that the -a, -A, -n and -N options treat '+' and '.' characters as normal alphabetic characters). The following description of perl regular expressions is mostly from the Perl Reference by Johan Vromans:



Each character matches itself, unless it is one of the special characters `^$+?.*O[]{}|`

- `^` at start of pattern, anchors pattern to the beginning of the string being matched.
- `$` at end of pattern, anchors pattern to the end of the string being matched.
- `.` matches any arbitrary character, but not a newline.
- `(...)` groups a series of pattern elements to a single element.
- `+` matches the preceding pattern element one or more times.
- `?` matches zero or one times.
- `*` matches zero or more times.
- `{N,M}` denotes the minimum N and maximum M match count. `{N}` means exactly N times; `{N,}` means at least N times.
- `[...]` denotes a class of characters to match. `[^...]` negates the class.  
Inside a class, `-` indicates a range of characters.
- `(...|...|...)` matches one of the alternatives.

Non-alphanumerics can be escaped from their special meaning using a backslash (`\`).  
Backslash is also used to form more special characters:

- `\w` matches alphanumeric, including ``_'`,
- `\W` matches non-alphanumeric.
- `\b` matches word boundaries,
- `\B` matches non-boundaries.
- `\s` matches whitespace,
- `\S` matches non-whitespace.
- `\d` matches numeric,
- `\D` matches non-numeric.

Examples: # Summarize:

```
wwwstat -a '.com$'          # only reqs from US commercial orgs.
wwwstat -a '^simplon.ics.uci.edu$' # only reqs from that hostname
wwwstat -A '^simplon.ics.uci.edu$' # no reqs from that hostname
wwwstat -c '302'            # only redirected requests
wwwstat -c '^5'             # only reqs resulting in server errors
wwwstat -C '200'            # only unsuccessful requests
wwwstat -d '[1-7]'          # only the first week of the month
wwwstat -d '([89]|1[0-4])'  # only the second week of the month
wwwstat -d '(1[5-9]|2[01])' # only the third week of the month
wwwstat -d '2[2-8]'         # only the fourth week of the month
wwwstat -d '(29|30|31)'     # only the leftover days of the month
wwwstat -d 'Feb'            # only February log entries
wwwstat -d '1994'           # only year 1994 log entries
wwwstat -D 'Apr'            # no entries from April
wwwstat -t '00'             # only reqs between midnight and 1am
wwwstat -T '12'             # no reqs between noon and 1pm
wwwstat -n '.gif$'          # only those reqs with a gif extension
wwwstat -n '^/~user/'        # only those reqs under user's directory
wwwstat -N '/hidden/'       # no reqs for files under "hidden" dirs
```

Depending on your unix shell, some special characters may need to be escaped on the command line to avoid shell interpretation.

The intention is that wwwstat be run by a wrapper program as a crontab entry at midnight, with its output redirected to a temporary file which can then be moved to the site's summary file. The temporary file is necessary to avoid erasing your published file during wwwstat's processing (which would look very odd if someone tried to GET it from your web). See below for example crontab entries.

One of the nicest things about wwwstat is that it does not make any changes to or write any files in the server directories. Thus, this program can be safely run by any user with read access to the httpd server's access\_log and srm.conf files. This allows people to do specialized summaries of just the things they are interested in.

This program could easily be modified to run as a CGI script, but that is not recommended for slow processors or heavily utilized servers unless some effort is made to keep the active log file very small (e.g. by using the -i option to bootstrap prior output of wwwstat).

---

### Frequently Asked Questions

#### 1. Why is all that legalese necessary? Isn't wwwstat free?

The above legalese exists because others have abused the privilege of using free software. Because this software was developed by an employee of the University of California, we must protect ourselves from lawsuits by those who would abuse our legal system for personal gain, regardless of any actual damages. To our knowledge, no damage has ever been caused by this program.

Wwwstat is distributed free of charge and will remain so as long as it is legally possible. If you are not distributing the program to others, there is no need for you to include mention of the University of California in its output. However, I would prefer that you leave in the reference to wwwstat's distribution site (at the bottom of the output) that others can know where to get the original program.

Wwwstat is in use around the world. If you have translated the output to another language (i.e. German, French, Maori, etc.), I encourage you to share those translations with others or mail them to me (Roy Fielding) so that I can provide special patch files for each language.

#### 2. Will you be developing a version for other httpd's, i.e. CERN, Plexus, ...

Obviously, versions of this program would also be nice for the Plexus and CERN servers. However, I found that much of the logic for finding file names was just too specific to the NCSA server to justify all the other work of making this general. Although this should now be easier given the common logfile format, I don't have the time to install all those servers just to see how to do it. Feel free to do so yourself.

#### 3. Why use a separate program (oldwwwstat) for prior log formats?

Why not just use a command-line option or examine the log content?

Because prior versions of wwwstat required a great deal of special file- handling capability to find file size information. Since that is no longer needed, it would be a waste to leave it in.

Eventually, all systems will migrate to the new format (or something like it) and having to maintain the old code without being able to test it is just plain silly.

#### 4. How do I setup a crontab script to run wwwstat nightly?

Well, that depends on how your system's crontab works, but on mine (a Sun 4 running SunOS 4.1.2) I can edit the crontab with the command `% crontab -e`

I have the following entry for my nightly script:

`0 0 * * * /dc/ud/www/etc/update-stats` and the following is my update stats script (thanks to Hal Varian)

```
-----
#!/bin/sh

/dc/ud/www/bin/wwwstat > /tmp/wwwstats.html
mv -f /tmp/wwwstats.html /dc/ud/www/documentroot/Admin/wwwstats.html
exit
-----
```

Here is another script submitted by LMD/T/AD Roar Smith:  
(NOTE I have not tried this myself, but it looks good).

```
---BEGIN wwwstat.cron-----
#!/bin/sh -fh
#
# wwwstat.cron
#
# Created: 1994-03-11 by LMD/T/AD Roar Smith (lmdrsm@lmd.ericsson.se)
# Modified: 1994-03-22 by LMD/T/AD Roar Smith (lmdrsm@lmd.ericsson.se)
# Wrote comments.
# Modified: 1994-04-05 by LMD/T/AD Roar Smith (lmdrsm@lmd.ericsson.se)
# Bug fix for first day of month.
#
# Copyright: This program is in the Public Domain.
#
#
# Run this script just after midnight on every day of the month.
#
# Example crontab entries:
# -----
# 1 0 * * * /library/WWW/wwwstat/wwwstat.cron
# -----
#
{
    program="/library/WWW/wwwstat/wwwstat-0.3/wwwstat"
    httpd="/usr/local/etc/httpd/httpd"
    statdir="/library/WWW/stats"
    statfile="wwwstats.html"
    tmpfile="/tmp/wwwstats.$$"
    accessfile="/var/adm/httpd_access_log"
```

```

errorfile="/var/adm/httpd_error_log"
pidfile="/var/adm/httpd.pid"

umask 022
day="/bin/date +%d"
month="/bin/date +%m"
set -- Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov
shift $month
lmonth="$1"
if [ "$day" -eq 1 ]; then
    #
    # First kill HTTP daemon to avoid interference
    #
    httpdpid="/bin/cat "$pidfile"
    if [ -z "$httpdpid" ]; then
        /bin/kill -TERM "$httpdpid"
    fi
    #
    # Copy Access and Error logfiles
    #
    cp -p "$accessfile" "$accessfile.$lmonth"
    cp -p "$errorfile" "$errorfile.$lmonth"
    /usr/etc/chown root.daemon "$accessfile.$lmonth"
    /usr/etc/chown root.daemon "$errorfile.$lmonth"
    #
    # Empty Access and Error logfiles
    #
    echo -n >"$accessfile"
    echo -n >"$errorfile"
    #
    # Restart HTTP daemon
    #
    (cd / ; "$httpd")
    #
    # Run stats program
    #
    $program -d "$lmonth" "$accessfile.$lmonth" >"$tmpfile" &&
    /bin/mv "$tmpfile" "$statdir/$lmonth.$statfile" &&
    /usr/etc/chown root.daemon "$statdir/$lmonth.$statfile"
    #
    # Copy this as current stats file
    /bin/cp -p "$statdir/$lmonth.$statfile" "$statdir/$statfile"
else
    #
    # Run stats program
    #
    $program >"$tmpfile" &&
    /bin/mv "$tmpfile" "$statdir/$statfile" &&
    /usr/etc/chown root.daemon "$statdir/$statfile"
fi

```

```

} 2>&1 |
mail webmaster 2>&1 1>/dev/null
exit
---END wwwstat.cron-----

```

#### 6. What is the general procedure for monthly resetting of the access\_log?

Again, that depends a great deal on how your site is set up and how frequent the accesses are to your server. My site gets about 15000 requests a month, so I just do the following at the beginning of each month (the example is for April):

```

% cp httpd/logs/access_log oldlogs/Mar_access_log
% vi oldlogs/Mar_access_log
    -- then delete all entries that are not from March
    or that are obviously corrupted.
% wwwstat -e oldlogs/Mar_access_log > /tmp/Mar.wwwstats.html
    -- this creates the full monthly summary for March and at the
    same time (the -e option) lists out any other corrupted entries
    that I may want to delete from the log.
% mv /tmp/Mar.wwwstats.html documentroot/Admin/Mar.wwwstats.html
    -- to publish the summary on my web.
% gzip -9 oldlogs/Mar_access_log
    -- use compress if you don't have gzip.
% cd httpd/logs
% mv access_log access_log.tmp
    -- if using a standalone type server, send a kill -1 to the
    httpd process so that it creates a new access_log. This is
    not necessary for inetd servers.
% vi access_log.tmp
    -- then delete all entries from March (should now be left with
    only April entries, since this is repeated monthly).
% cat access_log >> access_log.tmp
% mv -f access_log.tmp access_log
    -- the above two commands should be done in quick succession to avoid missing
    a new entry, and then followed by a kill -1 to the httpd process if running in standalone.

```

#### 7. My server load is HUGE and wwwstat runs out of memory, what can I do?

The only solution I can recommend is to use the -i option and bootstrap wwwstat's output every day -- setup a process which purges the logfile every night and creates a wwwstat output file which can be included the next day, and so on. The process would do something like:

```

% mv -f httpd/logs/access_log /tmp/access_log
    if server is standalone, restart it with
    kill -1 `cat httpd/logs/httpd.pid`
% wwwstat -i docroot/stats/current.html /tmp/access_log > /tmp/wwwwout
% mv -f docroot/stats/current.html docroot/stats/previous.html
% mv -f /tmp/wwwwout docroot/stats/current.html
% cat /tmp/access_log >> archived_log
% rm -f /tmp/access_log

```

---

---

If you have any suggestions, bug reports, fixes, or enhancements, send them to the author Roy Fielding at <fielding@ics.uci.edu>. See the file Changes for known problems and complete version information.

This work has been sponsored in part by the Advanced Research Projects Agency under Grant Number MDA972-91-J-1010. This software does not necessarily reflect the position or policy of the U.S. Government and no official endorsement should be inferred. Their support is appreciated.

**2.4.- ANNEXES BASE DE DONNEES ACCESS**

**2.4.1.- Définition des Tables pour la Base de Données ACCESS**

**a) - TABLE ACCESSES:**

CHAMP	TYPE DE DONNEES
SitelD	Numérique
ObjectD	Numérique
AuthUserID	Numérique
TapUser	Date/heure
Method	Texte
Status	Numérique
Bytes	Numérique
Index Query	Oui/Non
Domain	Texte

**b)\_ TABLE OBJECT**

CHAMP	TYPE DE DONNES
ObjectID	Compteur
Object Name	Texte
Restrict	Numérique

**c) - TABLE SITES**

CHAMP	TYPE DE DONNEES
SitelD	Compteur
Site Name	Texte
Restrict	Numérique

**d) - TABLE PAST - TOTALS**

CHAMP	TYPE DE DONNEES
Report Number	Compteur
ReportStartDate	Date/heure
Accesses	Numérique
Bytes	Numérique
IndexQueries	Numérique
Gets	Numérique
Posts	Numérique
Heds	Numérique

**e) - TABLE AuthUser**

CHAMP	TYPE DE DONNEES
AuthUserID	Compteur
AuthUser Name	Texte
Restrict	Numérique

**f) - TABLE CurACC**

CHAMP	TYPE DE DONNEES
SiteID	Numérique
ObjectID	Numérique
AuthUserID	Numérique
TapUser	Texte
Time	Date/heure
Method	Texte
Status	Numérique
Bytes	Numérique
IndexQuery	Qui/Non
Domain	Texte



**g) - TABLE KEY**

CHAMP	TYPE DE DONNEES
Table	Numérique
Name	Texte

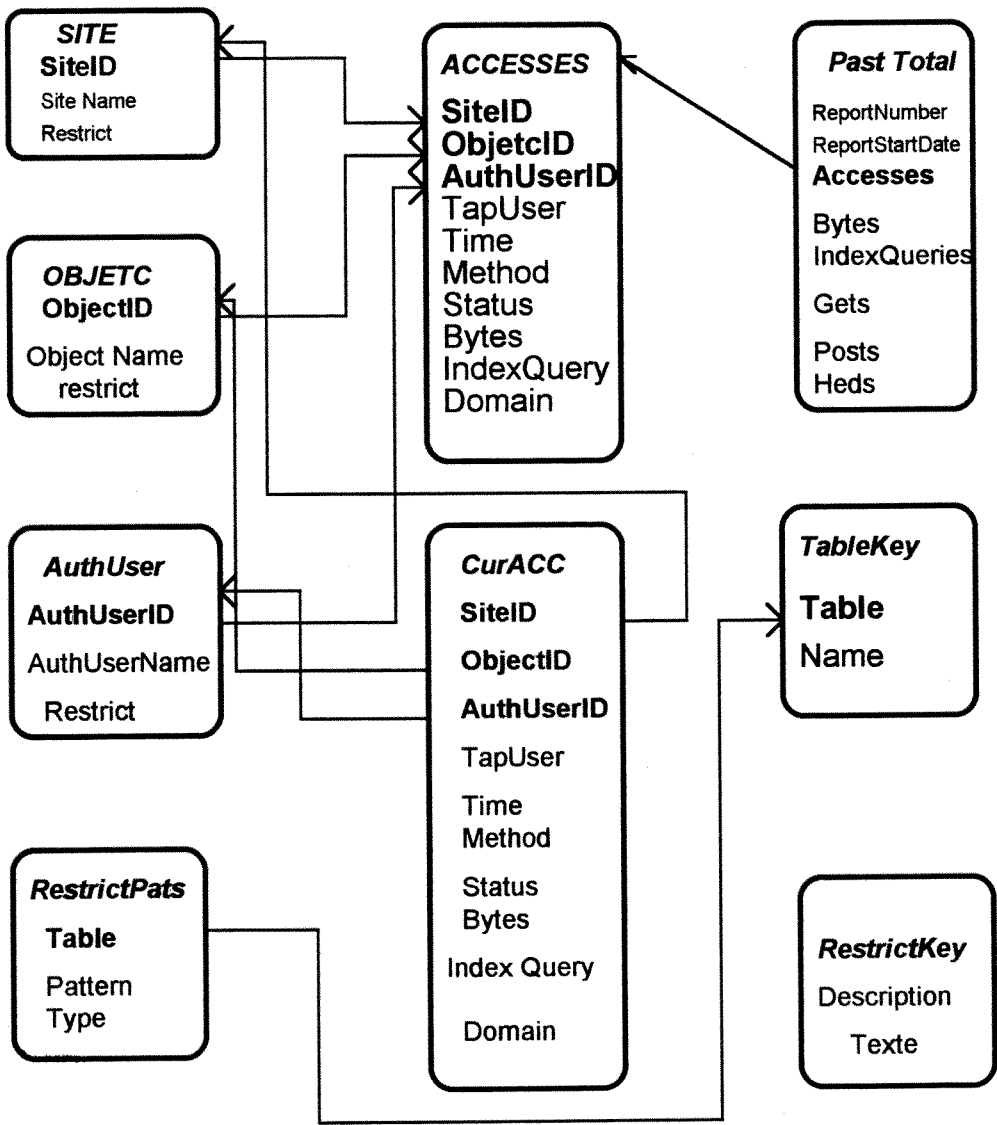
**h) - TABLE RestrictKey**

CHAMP	TYPE DE DONNEES
Restrict	Numérique
Description	Texte

**i) - TABLE RestrictPats**

CHAMP	TYPE DE DONNEES
Pattern	Texte
Table	Numérique
Type	Numérique

2.4.2.- RELATIONS DE LA BASE DE DONNEES



### 2.4.3.- REQUETES DEFINIES :

#### a) Comparaison des sites accédants Français et non Français

```
SELECT DISTINCTROW [00 Site Français accédant par jour].[Site Français], [00 site non Français par jour].[Site non Français], [00 Site Français accédant par jour].Temps  
FROM [00 Site Français accédant par jour] LEFT JOIN [00 site non Français par jour]  
ON [00 Site Français accédant par jour].Temps = [00 site non Français par jour].Temps  
GROUP BY [00 Site Français accédant par jour].[Site Français], [00 site non Français par jour].[Site non Français], [00 Site Français accédant par jour].Temps  
ORDER BY [00 Site Français accédant par jour].Temps;
```

#### b) Evolution de l'utilisation des Pages Flash Info

```
SELECT DISTINCTROW Sites.SiteName, Objects.ObjectID AS [Pages flash],  
Accesses.Time  
FROM (Accesses INNER JOIN Sites ON Accesses.SiteID = Sites.SiteID) INNER JOIN  
Objects ON Accesses.ObjectID = Objects.ObjectID  
WHERE ((Sites.SiteName Not Like "handy*" And Sites.SiteName Not Like "coms1.univ-lyon1.fr" And Sites.SiteName Not Like "coms2.univ-lyon1.fr" And Sites.SiteName Not Like "pixies*" And Sites.SiteName Not Like "p-girard.univ-lyon1.fr" And Sites.SiteName Not Like "l-de.univ-lyon1.fr") AND (Objects.ObjectName Like "**flash*"))  
GROUP BY Sites.SiteName, Objects.ObjectID, Accesses.Time  
ORDER BY Accesses.Time;
```

#### c) Nombre des Fichiers Consultables

```
SELECT DISTINCTROW Objects.ObjectName, Accesses.Time  
FROM (Objects INNER JOIN Accesses ON Objects.ObjectID = Accesses.ObjectID) INNER JOIN  
Sites ON Accesses.SiteID = Sites.SiteID  
WHERE ((Objects.ObjectName Not Like "**gif*" And Objects.ObjectName Not Like "**.gif" And Objects.ObjectName Not Like "*.xbm") AND (Sites.SiteName Not Like "handy*" And Sites.SiteName Not Like "p-girard.uni-lyon1.fr" And Sites.SiteName Not Like "coms*" And Sites.SiteName Not Like "pixies*" And Sites.SiteName Not Like "l-de.univ-lyon1.fr"))  
GROUP BY Objects.ObjectName, Accesses.Time;
```

#### d) Objet et nombre de fois consulté

```
SELECT DISTINCTROW Objects.ObjectName, Count(Objects.ObjectName) AS  
CompteDeObjectName  
FROM (Accesses INNER JOIN Objects ON Accesses.ObjectID = Objects.ObjectID) INNER JOIN  
Sites ON Accesses.SiteID = Sites.SiteID  
WHERE ((Objects.ObjectName Not Like "**gif/*" And Objects.ObjectName Not Like "**.gif" And Objects.ObjectName Not Like "*.xbm") AND (Sites.SiteName Not Like "handy*.univ-
```

lyon1.fr" And Sites.SiteName Not Like "pixies.univ-lyon1.fr" And Sites.SiteName Not Like "coms1.univ-lyon1.fr" And Sites.SiteName Not Like "coms2.univ-lyon1.fr" And Sites.SiteName Not Like "p-girard.univ-lyon1.fr" And Sites.SiteName Not Like "l-de.univ-lyon1.fr"))

**GROUP BY** Objects.ObjectName

**ORDER BY** Count(Objects.ObjectName) DESC;

### **e) Site Français accédant par jour**

**SELECT DISTINCTROW** Count(Sites.SiteName) AS [Site Français], Int([Time]) AS

Temps**FROM** Accesses **INNER JOIN** Sites **ON** Accesses.SiteID = Sites.SiteID

**WHERE** ((Sites.SiteName Like "\*.fr" And Sites.SiteName Not Like "handy\*.univ-lyon1.fr" And Sites.SiteName Not Like "p-girard.univ-lyon1.fr" And Sites.SiteName Not Like "l-de.univ-lyon1.fr" And Sites.SiteName <> "coms?.univ-lyon1.fr" And Sites.SiteName <> "pixies.univ-lyon1.fr"))

**GROUP BY** Int([Time])

**ORDER BY** Int([Time]);

### **f) Site non Français par jour**

**SELECT DISTINCTROW** Count(Sites.SiteName) AS [Site non Français], Int([Time]) AS

Temps**FROM** Accesses **INNER JOIN** Sites **ON** Accesses.SiteID = Sites.SiteID

**WHERE** ((Sites.SiteName Not Like "\*.fr"))

**GROUP BY** Int([Time]);

### **g) Site par jour**

**SELECT DISTINCTROW** Count(Accesses.SiteID) AS [Nombre de sites], Int([Time]) AS  
Expr1**FROM** Accesses

**GROUP BY** Int([Time]);

### **h) Sites et Objets par Mois**

**SELECT DISTINCTROW** Accesses.SiteID, Accesses.AuthUserID, Accesses.Time**FROM**  
Accesses

**WHERE** ((Accesses.SiteID In (SELECT [SiteID] FROM [Accesses] As Tmp

**GROUP BY** [SiteID],[AuthUserID]

**HAVING** Count(\*)>1 And [AuthUserID] = [Accesses].[AuthUserID]))

**ORDER BY** Accesses.SiteID, Accesses.AuthUserID;

### **i) Sites Français accédant à objet Cours**

```
SELECT DISTINCTROW Sites.SiteName, Accesses.ObjectID, Accesses.Time
FROM [00 Site Français accédant par jour], [00 Site Français accédant par jour] AS [00 Site
Français accédant par jour_1], [00 Site Français accédant par jour] AS [00 Site Français
accédant par jour_2], (Accesses INNER JOIN Objects ON Accesses.ObjectID =
Objects.ObjectID) INNER JOIN Sites ON Accesses.SiteID = Sites.SiteID
WHERE ((Sites.SiteName Like "*.fr") AND (Objects.ObjectName Like "*cours" And
Objects.ObjectName Like "cours"))
GROUP BY Sites.SiteName, Accesses.ObjectID, Accesses.Time
ORDER BY Sites.SiteName, Accesses.Time;
```

### **j) Sites non Français accédant à objet Cours**

```
SELECT DISTINCTROW Sites.SiteName, Objects.ObjectName, Int([time]) AS
TempsFROM (Accesses INNER JOIN Sites ON Accesses.SiteID = Sites.SiteID) INNER
JOIN Objects ON Accesses.ObjectID = Objects.ObjectID
WHERE ((Sites.SiteName Not Like "*.fr") AND (Objects.ObjectName Like "*cours"))
GROUP BY Sites.SiteName, Objects.ObjectName, Int([time])ORDER BY Sites.SiteName,
Int([time]);
```

### **k) Total Accesses**

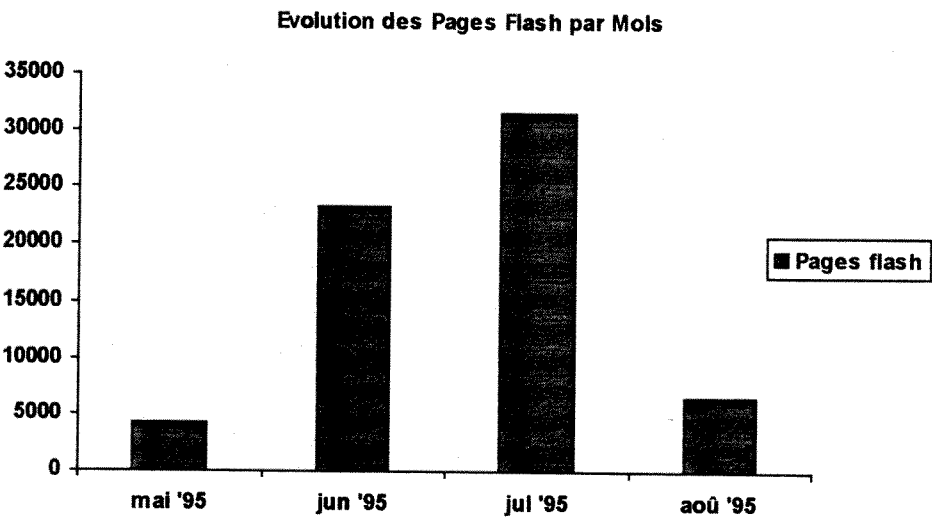
```
SELECT DISTINCTROW Count(*) AS Total
FROM (( CurAcc INNER JOIN Sites ON CurAcc.SiteID = Sites.SiteID) INNER JOIN
Objects ON CurAcc.ObjectID = Objects.ObjectID) INNER JOIN AuthUsers ON
CurAcc.AuthUserID = AuthUsers.AuthUserID
WHERE ((Objects.Restrict<2) AND (Sites.Restrict<2) AND (AuthUsers.Restrict<2));
```

### **l) Trouver les Doublons pour Accesses**

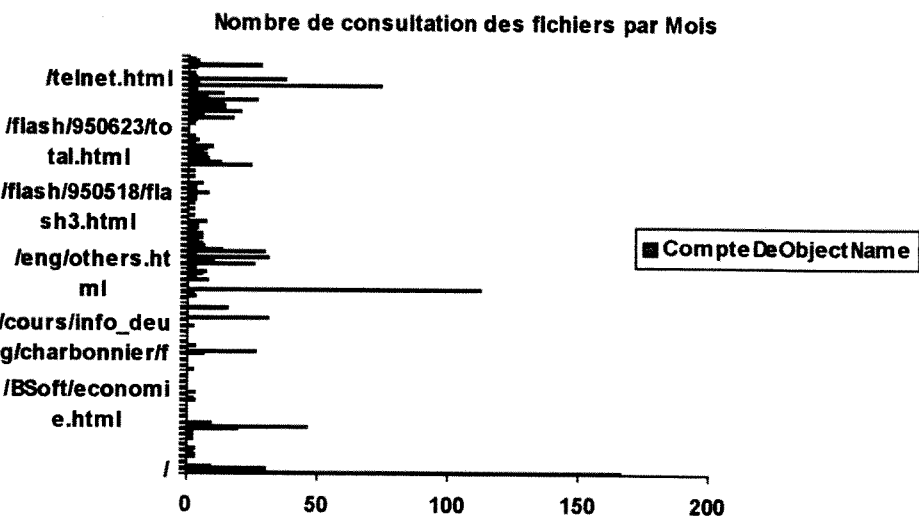
```
SELECT DISTINCTROW Accesses.ObjectID, Accesses.SiteID, Accesses.Time
FROM Accesses
WHERE ((Accesses.ObjectID In (SELECT [ObjectID] FROM [Accesses] As Tmp
GROUP BY [ObjectID] HAVING Count(*)>1 )))
ORDER BY Accesses.ObjectID;
```

2.4.4.- - GRAPHIQUES ET TABLEAUX

a)



b)



c)

## Objets plus consultés

Objet	Temps
/	888
/cours/sommaire.html	222
/eng/home.html	174
/services_offerts.html	170
/monde.html	142
/cgi-bin/wwwwais	139
/home.html	138
/autres.html	123
/cgi-bin/SFgate	93
/cours/mycot/mycot.html	81
/wais_interro.html	79
/flash/flash.html	77
/flag/couleur/uk.xpm	77
/condorcet/sommaire.html	76
/telnet.html	67
/materiels.html	66
/bibnv.html	66
/cours/info_deug/charbonnier/fiche_pascal/sommaire.htm	63
/new_etud.html	58
/cours/lamot/cours.c.html	51
/avh/avh_home_cata.html	49
/vienwin.html	49
/ch.html	48
/cours/info_deug/deug.int/semestr.1/sommaire.html	47
/lcons/network.gros.xbm	47
/decoret.html	47
/flash/950505/flash.html	45
/cgi-bin/wais.pl	44
/cours/info_deug/charbonnier/fiche_pascal/fiche1.html	37
/flash/	36
/equipbib.html	35
/condorcet/05alma.html	34
/prnote.html	34
/cgi-bin/SFgate2	33
/cgi-bin/waismail	31
/bibnew.html	31
/mh_geog_ok.html	29
/cours/info_deug/deug.int/semestr.1/annea.html	29
/flash/950623/flash.html	28
/usage/	28
/examen.html	28
/cours/info_deug/sommaire.html	27
/cours/info_deug/charbonnier/fiche_pascal/fiche2.html	26
/cours/info_deug/deug.int/semestr.1/calcs.html	25
/eng/provided_service.html	23
/stats/index.html	23
/flash/950518/flash.html	23
/usage/index.html	22

/books.html	21
/condorcet/01-pr.html	20
BSoft/notesbs.html	19
/flash/950505/flash01.html	18
/newadr.html	17
/condorcet/00-me.html	17
/formul.html	16
/eng/world.html	15
/flash/950623/flash01.html	15
/condorcet/02bib.html	14
/flash/flash.htm	14
/flash/950505/flash02.html	14
/getstats/	14
/flash/950505/flash04.html	14
/icons/blank.xbm	14
/flash/950606/flash.html	14
/icons/back.xbm	14
/flash/950505/total.html	14
/usr/local/etc/httpd/cgi-bin/SFgate2	13
/usage/week22.html	13
/icons/text.xbm	13
/tiers.html	13
/cours/info_deug/charbonnier/fiche_pascal/fiche6.html	12
/flash/950623/total.html	12
BSoft/table.html	12
/cours/info_deug/fich0.html	12
/condorcet/06tapr.html	12
/usr/local/etc/httpd/cgi-bin/SFgate	12
/eng/new_etud.html	12
/cours/info_deug/charbonnier/fiche_pascal/fiche4.html	11
/cours/lamot/node1.html	11
/cours/info_deug/charbonnier/fiche_pascal/fiche5.html	11
/condorcet/12mesm.html	11
/icons/unknown.xbm	10
/condorcet/07taxt.html	10
/avh/blocno1.html	10
/condorcet/04alm.html	10
/flash/950505/flash03.html	10
/flash/950623/flash05.html	9
/cours/lamot/node3.html	9
/flash/950518/total.html	9
/eng/others.html	9
/flash/950505/total.ps	9
/usage	9
/flash/950623/flash02.html	9
/wwwstat-1.0/Jan.wwwstats.html	8
/wusage/index.html	8
/condorcet/13tabl.html	8
/cours/info_deug/charbonnier/fiche_pascal/fiche3.html	8
/condorcet/11mesm.html	8
/condorcet/09man.html	8
/flash/950623/flash03.html	8
/condorcet/03bio.html	8
/cours/info_deug/deug.int/semestr.1/equa.html	8
/vacat.html	8
/cours/lamot/node4.html	8
/cours/info_deug/fich1.html	8



/cours/lamot/node6.html	8
/flash/950505/flash05.html	7
/flash/950725/flash.html	7
/cours/info_deug/deug.int/semestr.1/c-fonct.html	7
/wwwstat-1.0/results.html	7
/cours/info_deug/deug.int/semestr.1/nombr.html	7
/cours/lamot/node2.html	7
/cours/lamot/node5.html	7
/wais_interro.html>Serveur	7
/cours/lamot/node7.html	7
/cours/info_deug/fich2.html	7
/icons/menu.xbm	7
/eng/eng_ch.html	6
/wwwstat-1.0/May.wwwstats.html	6
/flash	6
/cours/info_deug/fich4.html	6
/flash/950518/total.ps	6
/avh/ecranb2.html	6
/BSoft/info.html	6
/stats/	6
/condorcet/05alma_fn.html	6
/flash/950606/flash3.html	6
/cours/info_deug/deug.int/semestr.1/reisr.html	6
/cours/info_deng/	6
/flash/950505/flash07.html	6
/flash/950505/flash06.html	6
/flash/950623/flash04.html	6
/eng/prnote.html	6
/avh/divers.html	5
/usage/week18.html	5
/flash/950518/flash2.html	5
/avh/imprim5.html	5
/cours/info_deug/deug.int/semestr.1/calcu.html	5
/stats	5
/flash/950725/total.html	5
/avh/logibur.html	5
/cours	5
/avh/imprim.html	5
/avh/scanner.html	5
/avh/divers2.html	5
/avh/ecranb.html	5
/usage/week23.html	5
/condorcet/10mand.html	5
/icons/image.xbm	5
/nsage/week1341.html	5
/condorcet/	5
/cgi-bin/statform	5
/work/format-ph.html	5
/condorcet/04alm_fn.html	5
/01-pr_fn.html	4
/flash/950518/flash3.html	4
/flash/950518/flash7.html	4
/flash/950518/flash6.html	4
/flash/flash01.htm	4
/cours/info_deug/deug.int/semestr.1/limit.html	4
/cours/info_deug/deug.int/semestr.1/group.html	4
/flash/flash01.html	4

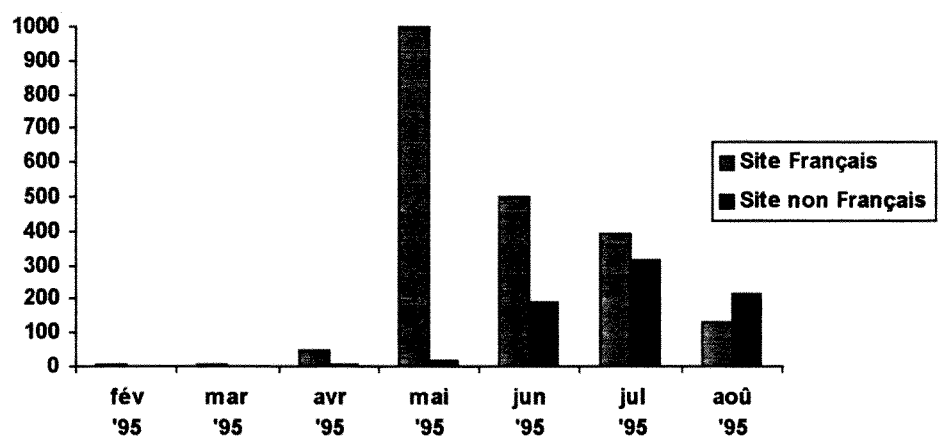
/cours/info_deug/deug.int/	4
/cours/lamot/node8.html	4
/avh/blocnoe.html	4
/flash/950725/flash01.html	4
/flash/950725/flash05.html	4
/cours/	4
/cours/info_deug/fich3.html	4
/cours/info_deug/deug.int/semestr.1	4
/BSoft/biograph.html	4
/BSoft/romans.html	4
/flash/950606/flash1.html	4
/flash/950606/flash2.html	4
/cgi-bin/essai	4
/flash/950606/flash5.html	4
/usage/week12.html	4
/BSoft/distrib.html	4
/BSoft/logiciel.html	3
/BSoft/pedagog.html	3
/BSoft/medecine.html	3
/condorcet/textes/new/	3
/BSoft/histoire.html	3
/BSoft/economie.html	3
/usage/week1342.html	3
/cours/lamot/node13.html	3
/BSoft/nouvelle.html	3
/cours/lamot/node11.html	3
/eng/andy.univ-lyon1.fr/cours/sommaire.html	3
/flash/950623/total.ps	3
/provided_service.html	3
/robots.txt	3
/flash/950518/flash1.html	3
/eng/andy.univ-lyon1.fr/eng/home.html	3
/archives_old	3
/.	3
/wwwstat-1.0/Apr.wwwstats.html	3
/cours/lamot/node19.html	3
/usage/week11.html	3
/cours/lamot/node128.html	3
/flash/950725/flash02.html	3
/usage/week21.html	3
/usage/week17.html	3
/flash/950725/flash03.html	3
/condorcet	3
/flash/950725/flash04.html	3
/usage/week8.html	3
/condorcet/textes/	3
/usage/week0.html	3
/flash/flash07.html	2
/flash/flash03.html	2
/flash/flash02.html	2
/flash/flash06.html	2
/usage/week10.html	2
/flash/flash04.html	2
/usage/week13.html	2
/flash/950606/flash4.html	2
/usage/week24.html	2
/index.html	2

/cours/info_deug/sommaire.html>TP	2
/wwwstat-1.0/Mar.wwwstats.html	2
/cours/info_deug/charbonnier/fiche_pascal/fiche0.html	2
/cgi-bin/wwwwais2	2
/wwwstat-1.0/Feb.wwwstats.html	2
/cgi-bin/print_hit_bold.pl/vienwin.html	2
/cours/info_deug/deug.int/semestr.1/fonc.html	2
/cgi-bin/print_hit_bold.pl/newadr.html	2
/usage/week1340.html	2
/BSoft/4ecouv.html	2
/BSoft/vieprat.html	2
/BSoft/theatre.html	2
/condorcet/05not.html	2
/BSoft/sciences.html	2
/BSoft/revues.html	2
/usage/week9.html	2
/cours.html	2
/flash/flash05.htm	2
/condorcet/textes	2
/BSoft/droit.html	2
/eng/examen.html	2
/usage/week4.html	2
/usage/week7.html	2
/BSoft/corresp.html	2
/BSoft/cdrom.html	2
/condorcet/01-pr_fn.html	2
/flash/950606/total.html	2
/BSoft/qsaisje.html	2
/usage/week20.html	2
/cours/lamot/node12.html	2
/cours/lamot/node14.html	2
/cours/lamot/node15.html	2
/cours/lamot/node16.html	2
/cours/info_deug/deug.int/semestr.1/	2
/flash/950518/flash4.html	2
/flash/950606/flash6.html	2
/export/home/deberti/Mail/projet_TeX	2
/condorcet/11mesm_fn.html	2
/avh/imprim3.html	2
/hh/newadr.html	2
/cours/lamot/node21.html	2
/cours/lamot/node20.html	2
/usage/week16.html	2
/cours/info_deug/charbonnier/fiche_pascal/sommaire.htm	2
/cours/lamot/node32.html	2
/cgi-bin/ACCES	2
/flash/950518/flash5.html	2
/cgi-bin/test-cgi	2
/ftp_index.html	2
/demo.html	2
/condorcet/mycotox.html	1
/avh/scanner.htm	1
/cours/lamot/node114.html	1
/usage/week1.html	1
/cgi-bin/print_hlt_bold.pl/telnet.html	1
/usage/week1339.html	1
/950505/flash.html	1

/usage/week14.html	1
/usage/week770.html	1
/usage/week905.html	1
/cours/lamot/node113.html	1
/condorcet/01-prefa.html	1
/usage/week15.html	1
/condorcet/tritraill/	1
/usage/week916.html	1
/eng/home.html%25	1
/usr/local/etc/httpd/htdocs/condorcet/13tabl.html	1
/cours/lamot/node43.html	1
/hh/vienwin.html	1
/hh/equipbib.html	1
/cours/info_deug/old_version/fich2.html	1
/cours/lamot/node116.html	1
/cours/lamot/node115.html	1
/condorcet/05ali.html	1
/BSoft/logiciels.html	1
/condorcet/00-merci.html	1
/formul2.html	1
/cgi	1
/mini>gif	1
/cgi-bin/test	1
/home.htm	1
/usage/week25.html	1
/usage/week19.html	1
/cours/lamot/node37.html	1
/cours/lamot/node52.html	1
/test_formul.html	1
/cours/lamot/node119.html	1
/nsage/week2.html	1
/cgi-bin/print_hit_bold.pl/eng/eng_ch.html	1
/usage/week3.html	1
/end/home.html	1
/decoret@handy.univ-lyon1.fr	1
/cours/lamot/node59.html	1
/flash/950725/total.ps	1
/fidevpr.html	1
/cgi-bin/Base	1
/cgi-bin/UCBL	1
/usage/week6.html	1
/usage/week5.html	1
/cgi-bin/wasi.pl	1
/usage/week27.html	1
/cours/info_deug/old_version/fich3.html	1
/cours/lamot/node42.html	1
/hh/decoret.html	1
/cours/mycot/	1
/eng/	1
/usr/local/etc/httpd/htdocs/eng/eng_ch.html	1
/flash/950606/total.ps	1
/cours/lamot/node10.html	1
/cours/lamot/node9.html	1
/flash/flash02.htm	1
/hh/home.html	1
/hh/ch.html	1
/condorcet/06tapr_fn.html	1

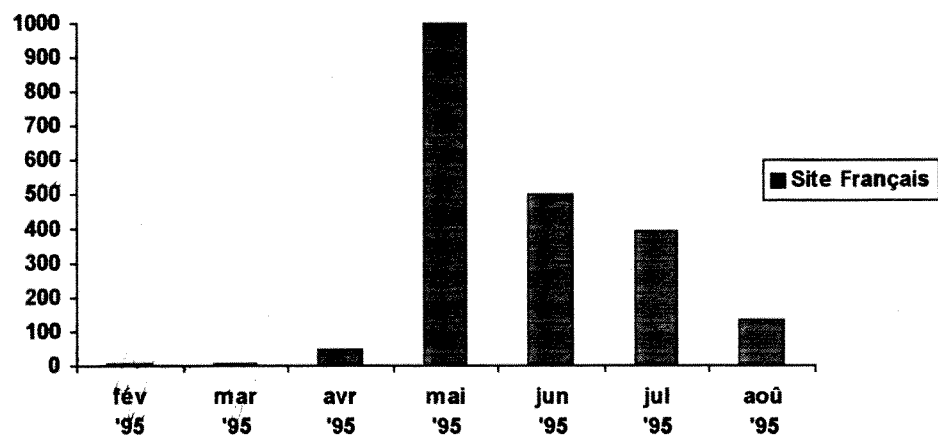
/cgi-bin/print_hit_bold.pl/ftp_index.html	1
/hh/buequipm.html	1
/condorcet/07taxt_fin.html	1
/deberti/cv.html	1
/cgi-bin/print_hit_bold.pl/eng/home.html	1
/deberti/cv.ht	1
/flash/flash05.html	1
/usr/local/etc/httpd/htbin/SFgate	1
/*	1
/eng/HTTP//	1
//cgi-bin/wwwwais	1
/usr/local/etc/httpd/htdocs/vienwin.html	1
/flash/950518/flash03.html	1
/gif	1
/cgi-bin/print_hit_bold.pl/bibnv.html	1
/archives_old/equipbib_old.html	1
/condorcet/textes/new/couv4.rtf	1
/condorcet/05note_f.html	1
/handynet.html	1
/hh/telnet.html	1
/cgi-bin/bibnv.html	1
/cours/info_deug/old_version/fich1.html	1
/cours/lamot/node122.html	1
/cours/info_deug/old_version/sommaire.html	1
/cours/info_deug/old_version/	1
/cours/lamot/node91.html	1
/buequipm.html	1
/eng	1
/provided_services.html	1
/conrs/..	1
/cgi-bin/fich1.html	1
/cours/info_deug/charbonnier/fiche_pascal/fich1.html	1
/BSoft/guides.html	1
/hh/ftp_index.html	1
/hh/materiels.html	1
/cours/lamot/node17.html	1
/cours/info_deug/charbonnier/fiche_pascal/fich0.html	1
/cgi-bin/print_hit_bold.pl/bibnew.html	1
/cours/info_deug/fich2.rtf	1
/condorcet/textes/new/titre.html	1
/cours/info_deug	1
/oio	1
/test2_formul.html	1

d) Site Français et Non français par Mois

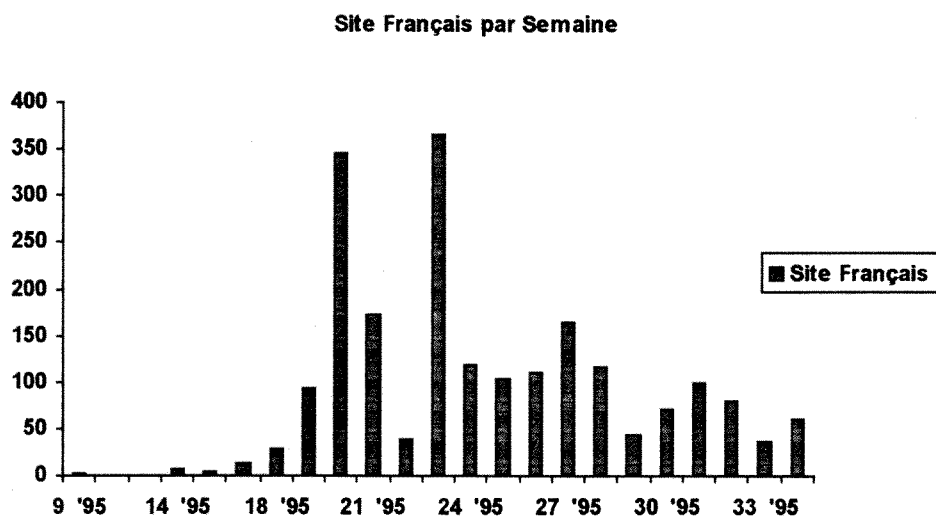


e)

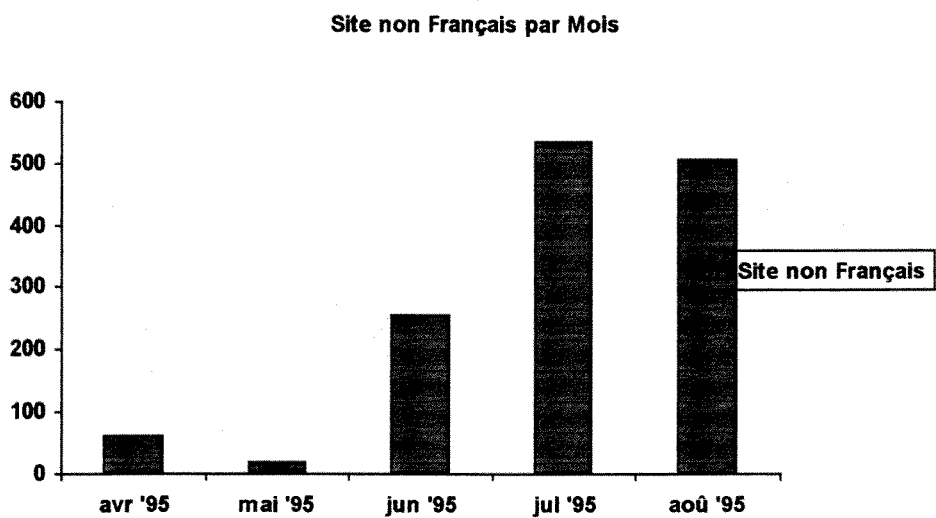
Site Français par mois



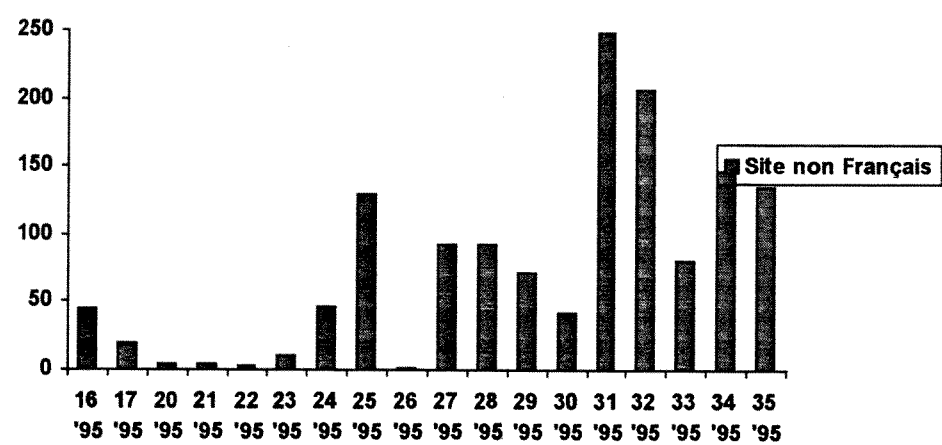
f)



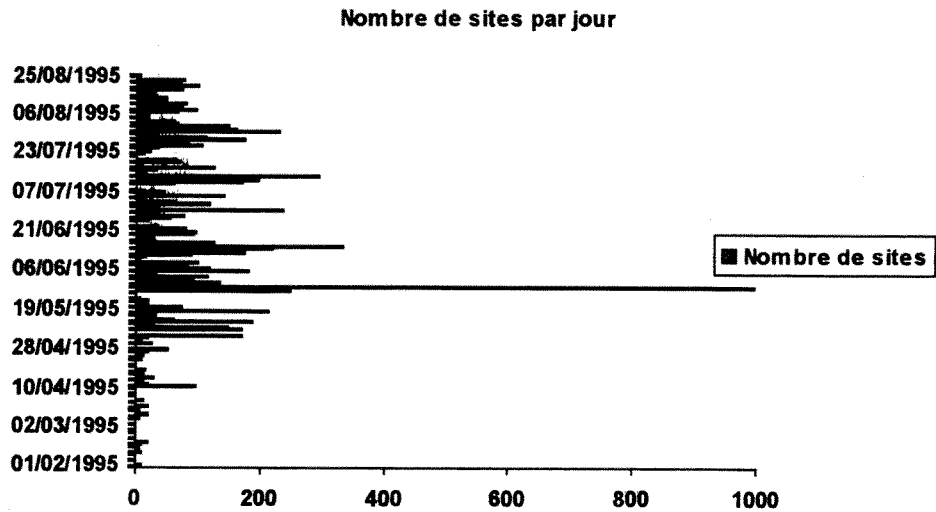
g)



**h) Site Non français par Semaine**

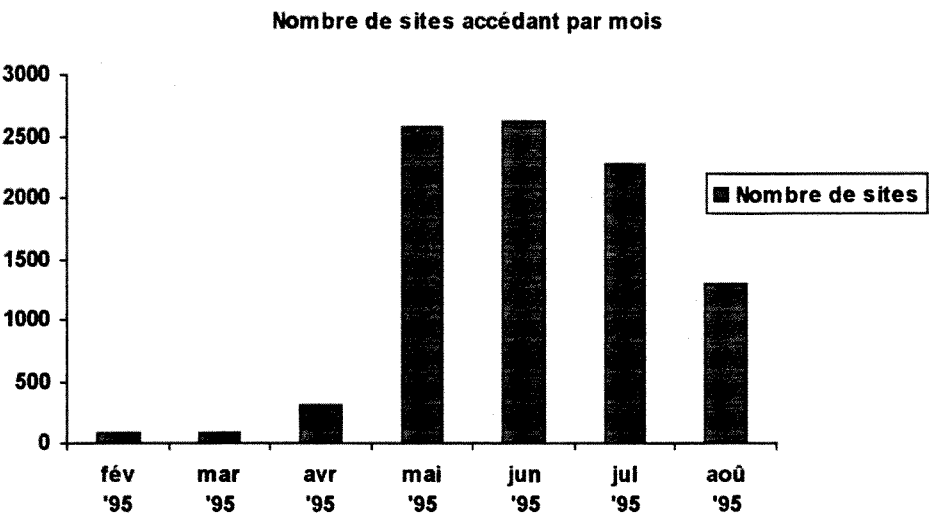


**i)**

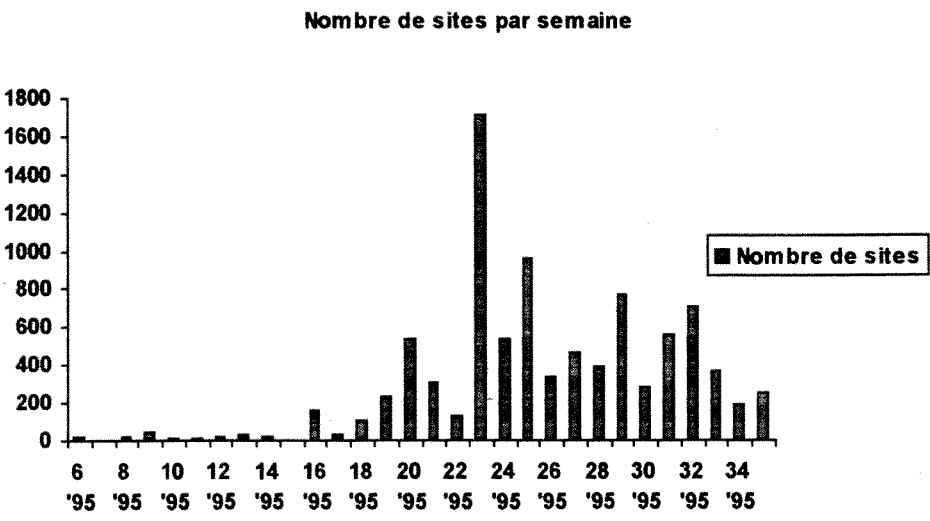




j)



k)



1)

## Objet et nombre de fois consulté

12-sep-95

ObjectName	CompteDeObjectName
/	166
/autres.html	30
/avh/avh_home_cata.html	9
/avh/blocno1.html	1
/avh/blocnoe.html	1
/avh/divers.html	3
/avh/divers2.html	3
/avh/ecranb.html	2
/avh/ecranb2.html	3
/avh/imprim.html	1
/avh/imprim3.html	1
/avh/imprim5.html	2
/avh/logibur.html	2
/avh/scanner.html	2
/bibnew.html	19
/bibnv.html	46
/books.html	9
/BSoft/4ecouv.html	1
/BSoft/cdrom.html	1
/BSoft/distrib.html	1
/BSoft/droit.html	1
/BSoft/economie.html	1
/BSoft/histoire.html	1
/BSoft/info.html	3
/BSoft/logiciel.html	2
/BSoft/medecine.html	1
/BSoft/notesbs.html	3
/BSoft/nouvelle.html	1
/BSoft/pedagog.html	1
/BSoft/qsaisje.html	1
/BSoft/revues.html	1
/BSoft/romans.html	1
/BSoft/sciences.html	1
/BSoft/table.html	2
/BSoft/theatre.html	1
/BSoft/vieprat.html	1
/cgi-bin/print_hit_bold.pl/b	1
/cgi-bin/SFgate	1
/cgi-bin/wwwwais	6
/ch.html	26
/condorcet/10mand.html	1
/condorcet/sommaire.html	3
/cours/info_deug/charbonnier	1
/cours/info_deug/deug.int/se	1
/cours/info_deug/deug.int/se	1
/cours/info_deug/deug.int/se	1
/cours/info_deug/sommaire.ht	1
/cours/lamot/cours.c.html	2
/cours/lamot/node1.html	1
/cours/mycot/mycot.html	1
/cours/sommaire.html	31
/deberti/cv.ht	1
/deberti/cv.html	1
/decoret.html	15
/decoret@handy.univ-lyon1.fr	1
/eng	1
/eng/	1
/eng/andy.univ-lyon1.fr/cour	3
/eng/examen.html	2
/eng/home.html	112
/eng/home.html%25	1
/eng/HTTP//	1
/eng/new_etud.html	8
/eng/others.html	3
/eng/pmote.html	5
/eng/provided_service.html	7

/eng/world.html	3
/equipbib.html	25
/examen.html	10
/flag/couleur/uk.xpm	31
/flash	3
/flash/	29
/flash/950505/flash.html	13
/flash/950505/flash01.html	6
/flash/950505/flash02.html	5
/flash/950505/flash03.html	4
/flash/950505/flash04.html	5
/flash/950505/flash05.html	5
/flash/950505/flash06.html	3
/flash/950505/flash07.html	4
/flash/950505/total.html	4
/flash/950518/flash.html	7
/flash/950518/flash1.html	1
/flash/950518/flash2.html	2
/flash/950518/flash3.html	1
/flash/950518/flash4.html	2
/flash/950518/flash5.html	1
/flash/950518/flash6.html	2
/flash/950518/flash7.html	3
/flash/950518/total.html	3
/flash/950606/flash.html	8
/flash/950606/flash1.html	3
/flash/950606/flash2.html	3
/flash/950606/flash3.html	5
/flash/950606/flash4.html	1
/flash/950606/flash5.html	2
/flash/950606/flash6.html	1
/flash/950606/total.html	2
/flash/950606/total.ps	1
/flash/950623/flash.html	24
/flash/950623/flash01.html	12
/flash/950623/flash02.html	8
/flash/950623/flash03.html	7
/flash/950623/flash04.html	5
/flash/950623/flash05.html	7
/flash/950623/total.html	9
/flash/950623/total.ps	2
/flash/950725/flash.html	4
/flash/950725/flash01.html	2
/flash/950725/flash02.html	1
/flash/950725/flash03.html	1
/flash/950725/flash04.html	1
/flash/950725/flash05.html	2
/flash/950725/total.html	3
/flash/flash.html	17
/formul.html	5
/home.html	20
/materiels.html	14
/mh_geog_ok.html	14
/monde.html	13
/new_etud.html	26
/newadr.html	7
/prnote.html	13
/robots.txt	3
/services_offerts.html	74
/stats/index.html	4
/telnet.html	37
/tiers.html	3
/usage/	2
/usage/week27.html	1
/vacat.html	3
/vienwin.html	28
/wais_interro.html	4
/work/format-ph.html	2

## **III.- ANNEXES CONVERTISSEURS**

### **3.1.- CYBERLEAF:**

Internet Publishing Made Simple

#### Benefits

- \* Open to popular word processors
  - \* Easy to use
  - \* Quality data conversion
  - \* Fast and productive document processing
  - \* Web lifecycle management
- 

#### Mastering the World Wide Web

With Cyberleaf, publishing information on the Internet's World Wide Web couldn't be easier or more practical. Use your favorite word processor to create documents as you normally would for any print application. Then use Cyberleaf to transform those very same documents into Internet webs, ready to be viewed by standard Web browsers, such as NCSA Mosaic, netscape Navigator and others.

#### The Cyberleaf advantage

There are many advantages to using Cyberleaf rather than specialized HTML authoring systems. Cyberleaf's push-button user interface requires no HTML expertise, so it's easy to use at any skill level.

Cyberleaf goes beyond HTML authoring tools, providing a complete web production environment that complements your existing word processor. Cyberleaf eliminates the need to use a separate, complex authoring tool for the World Wide Web. It enables you to maintain your information in a single source, so it can be easily managed and reused.

Through a simple conversion and web construction process, Cyberleaf turns the output of major word processing applications into Internet- ready webs that can be published on both private corporate networks and the public Internet's WorldWide Web.

#### Comprehensive web production

Cyberleaf provides the easiest, most comprehensive web production environment available, including the following functionality:

- \* Conversion of text, graphics and tables from Word RTF, WordPerfect, Interleaf and FrameMaker formats to HTML and GIF
- \* Intelligent HTML style matching
- \* Thumbnail graphics creation
- \* Hyperlinking between and among documents and other file types, including multimedia
- \* Home page templates that can be customized for unique applications
- \* Web lifecycle management, including incremental updates and link management

Système innovant de Gestion Electronique de Documents reconnu, DORIS est aussi un puissant logiciel documentaire. Sa fiabilité est garantie par le choix des normes internationales :

- . UNIX pour le système d'exploitation
- . SQL pour le stockage des données
- . ASN 1 (Abstract Syntax Notation 1) pour la présentation des données
- . X-WINDOW pour l'interface utilisateur

DORIS est caractérisé par une grande souplesse d'utilisation et d'adaptation. C'est un outil parfaitement adapté aux besoins spécifiques des utilisateurs, quel que soit le volume de données à gérer. Il peut répondre à des applications très diverses de type Gestion Documentaire mais aussi et surtout Gestion Electronique de Documents.

Enfin, le système DORIS est rendu vivant par la programmation orientée objet, technique de programmation bien plus intuitive que la programmation traditionnelle : DORIS est écrit en langage EIFFEL et en langage C.

EIFFEL est un langage Orienté Objet, dont la philosophie de "classes" et d'interaction de classes est très proche de la philosophie documentaire classique d'organisation de l'information et de ses traitements. Les programmes écrits en langage Objet sont reconnus comme étant réutilisables, extensibles, fiables et portables.

C'est à partir du moteur documentaire DORIS que le développement de LORIS, logiciel de Gestion de Bibliothèque, a été envisagé. LORIS, qui est encore à l'état de maquette, permet la gestion de toutes les fonctions bibliothéconomiques souhaitables, rendues performantes par la puissance documentaire et la gestion électronique de documents propres à DORIS.

Cyberleaf includes everything you need to transform information into Internet webs. With Cyberleaf, mastering the World Wide Web is just a mouse click away.

Cyberleaf provides:

1. Openness

Transforms documents created with your favorite word processor into Internet-compatible webs of information.

2. Ease of use

Simple, point-and-click user interface is ideal for both technical and non-technical users.

3. Quality

High-quality data conversion ensures data integrity.

4. Productivity

Single pass conversion of text, graphics and tables makes document processing fast and easy.

5. Lifecycle

Internet webs are easy to update as new files are added or existing ones modified.

---

"To date the Internet has largely been the haven of academia and has been vastly underutilized by corporations. Cyberleaf helps organizations mount the technology hurdle of putting information on the Internet, making the formerly complex become simple."

Linda Myers-Tierney, Senior Consultant, Workgroup Applications Systems, IDC

---

Overview

Cyberleaf allows you to create hyperlinked text and graphic documents, ready for distribution over the Internet, from documents you've created using your regular word processing software. Whether you use Word, WordPerfect, Interleaf, FrameMaker or any combination, Cyberleaf will enable you to create and maintain professional-looking Internet webs of information, which can be distributed on the World Wide Web or private corporate networks.

Simple web construction

Cyberleaf's unique, workflow-oriented interface guides you through the process of constructing a hyperlinked Internet web.

- \* Simply select your source files and push a button to start the conversion and web construction process.
- \* Use the batch processing option to collectively process multiple documents.
- \* Choose between transparent or white graphic backgrounds for your processed documents.
- \* Graphics can reside within documents or can be externally referenced through hyperlinks.
- \* Use the preview option to examine the results of each stage of web construction, including conversion, linking and the final web.

## Document conversion

Cyberleaf automatically converts popular word processing formats to HTML (HyperText Markup Language) and GIF (Graphics Interchange Format), the de facto standards for the Internet and World Wide Web.

- \* Converts Word RTF, WordPerfect, Interleaf, FrameMaker and ASCII formats to HTML and GIF.
- \* Simultaneously converts text to HTML and graphics to GIF or PostScript.
- \* Converts native Interleaf and FrameMaker hyperlinks directly to HTML.
- \* Converts tables to GIF inline raster images, PostScript or experimental HTML 3.0 table markup. (HTML 3.0 table support has been provided in anticipation of HTML 3.0 approval.)
- \* Lets you define your own document attributes and properties to guide you through the conversion process.
- \* Preserves footnotes automatically adding HTML hyperlinks to corresponding information.
- \* Cyberleaf output has been tested with popular Internet browsers such as NCSA Mosaic, Enhanced NCSA Mosaic from Spyglass, Cello, Spry AIR Mosaic, Netscape Navigator and Lynx.

## Style matching

During the conversion process, Cyberleaf analyzes the styles contained in the original documents and automatically matches those styles to equivalent HTML styles.

- \* Simple names for HTML styles eliminate the need to understand complex HTML syntax.
- \* You can easily modify default style matching through a point-and-click interface, allowing you to fine tune the look of your web.

## Thumbnail graphics

To save valuable time when downloading images within files, any graphic may be transformed into a thumbnail, a postage stamp-sized version of the original graphic.

- \* Thumbnail graphics are automatically hyperlinked to larger external images.
- \* End users who are viewing the information with Web browsers can simply click on the thumbnail to download the full-size graphic.

## Hyperlinking

An easy-to-use, point-and-click interface lets you insert hyperlinks within and between documents to form a cohesive web of information.

- \* Hyperlinks can be anchored to document names, text strings or text styles.
- \* Lets you hyperlink to HTML, GIF, PostScript, MPEG, JPEG, EPS, EPSI, TIFF and other file types.

- \* As source documents change, hyperlinks between files remain persistent and can simply be "reapplied" as the web is updated, preventing broken links.
- \* Lets you incorporate multimedia into your webs by creating hyperlinks to external programs, such as audio and video, which can then be launched from your website using a Web browser.
- \* Automatically generates hyperlink anchors for external graphics.
- \* Lets you hyperlink to URL references located at other sites.

## Home page creation

Professionally-designed templates provide a solid foundation for constructing effective home pages.

- \* Home page templates include attractive graphics, icons and hyperlink anchors for information distribution applications such as company information, customer support, government agency, documentation, multi-site project management and personal webs.
- \* Use the integrated home page editor to create personalized home pages for your web.

## Web document management

Only Cyberleaf makes it easy to manage the evolution of your webs. As documents change or new ones are added, simply pull out the old and plug in the new. Cyberleaf automatically ensures the integrity of your web.

- \* Lets you update your webs incrementally when your source documents change or new ones are added, eliminating the need to rebuild webs from scratch.
  - \* You can interactively refine and save the parameters used to convert your documents to facilitate quick updates and reuse throughout the lifetime of your web.
  - \* When documents are deleted or revised, Cyberleaf automatically reapplies previously defined styles and hyperlinks to the updated web, and identifies any broken links that require manual intervention.
  - \* To facilitate browsing, long documents may be outlined based on selected style names such as chapter or section, resembling a table of contents of your web.
  - \* Automatically inserts hyperlinks between document outline elements and their corresponding sections in the HTML output file for easy navigation.
  - \* Post web function automatically copies completed webs to your webserver.
  - \* Relative pathnames for web files simplify the movement of webs between file system locations.
- 

## About Interleaf

Interleaf is the world's leading provider of integrated document management solutions. Our hallmark is delivering open, scalable technology combined with world-class services and award-winning support. We have over 12 years experience implementing document management systems and over a million users worldwide.

Cyberleaf can be purchased separately or as a component of Interleaf's integrated suite of document management solutions.

---



For more information, please contact:

Interleaf Direct at 800-955-5323

email: [i-direct@ileaf.com](mailto:i-direct@ileaf.com)

your local Interleaf office

or Value-Added Reseller (VAR)

Copyright 1995 by Interleaf, Inc. All rights reserved. Interleaf is registered trademark of Interleaf, Inc. Cyberleaf is trademark of Interleaf, Inc. Other products and companies referenced herein are trademarks or registered trademarks of their respective companies or mark holders.

All Interleaf information presented on the Internet has been published using Interleaf's internet publishing software - Cyberleaf

Specifications subject to change without notice. (Rev. 4-95)

---

### 3.2.- RFTTOHTLM:

rtfhtml - A Filter to Translate RTF to HTML

RTF is a document formatting language developed by Microsoft. There are many word processing packages (Microsoft Word, Word Perfect, Frame Maker...) that can read or write RTF format files. rtfhtml was designed to translate existing RTF documents into HTML - the format of the World Wide Web. rtfhtml can also be used to author new documents specifically for the WWW.

The latest version of rtfhtml is 2.7.5

- \* rtfhtml Users Guide
- \* Frequently Asked Questions
- \* What's new this version
- \* Known Bugs and requested features.

How to Get the Package

rtf-ctrl has been Revised

When Word 6.0 was released, it introduced some new control words which rtfhtml 2.7.5 was not aware of. These caused messages to appear like:

ReadStyleSheet: unknown token "\widctlpar"  
Unknown symbol "\sprstsm" near line 9, position 66.  
Unknown symbol "\truncex" near line 9, position 74.  
Unknown symbol "\nolead" near line 9, position 81.  
Unknown symbol "\msmcap" near line 9, position 88.  
Unknown symbol "\widctlpar" near line 12, position 17.

In addition, Word 6.0 uses \super for footnotes instead of \up, which confuses the filter.

on 8/7/95 rtf-ctrl was again modified so that it correctly dealt with documents that contain revision marks.

To correct these problems, get the new version of rtf-ctrl and copy it into your folder or directory containing the filter.

Download the new rtf-ctrl

- \* Macintosh binary
- \* Source Package for UNIX
- \* sun4 binary
- \* sun4 Solaris binary
- \* sun4 Solaris binary
- \* OS2 binary
- \* DOS binary

Sample Documents

To get you started, I have a Style Sheet for Microsoft Word. This has definitions of the Paragraph styles that I use for authoring Web documents.  
It renders my RTF versions of the documents in a style that is close to what XMosaic 2.1 displayed (at one time.)

It is located here.

The HTML version is here.

#### About RTF

RTF (Rich Text Format) was designed by Microsoft as an open format for interchanging documents between Microsoft Word and other word processing packages. It is supported by WordPerfect, FrameMaker, Interleaf and many other packages on UNIX, Apple, Macintosh, Next, and PC platforms.

The specification for RTF and a set of filters for translating RTF to plain-text, troff and Tex and LaTeX are available at: <ftp://ftp.prima.wisc.edu/pub/RTF>

- This document was created from RTF source by rtfhtml version 2.7.5

--Graphics and rtfhtml

rtfhtml will generate a separate file for each picture found in your RTF source. The type of the picture file depends on how it was added to the RTF. If your word processor runs on Microsoft Windows, you will probably get a WMF (Windows Metafile Format) graphic. If your word processor runs on a Macintosh, you will get a PICT file. Since neither WMF nor PICT graphics are supported formats on the WWW, you will want to convert these files to GIF format. The following programs will convert to GIF format.

Name	Platform	Input formats
netpbm	UNIX	PICT
HiJaak Pro	Windows	WMF
Graphic Workshop	Windows	WMF
wmf2bmp	Windows	WMF
Graphic Converter	Mac	PICT and WMF
GIF Converter	Mac	PICT
Paint Shop Pro	Windows	WMF

Inset Systems, 71 Commerce Drive, Brookfield CT 06804-3405, Phone 203-740-2400, Fax 203-775-5634 They also have a toll free number 800-374-6738 (800 DR INSET). They have a BBS, 203-740-0063 and are on CompuServe (GO INSET)

The program runs under Microsoft Windows and says it can

- View, edit, and image-process a file
- Convert a file from one format to another (over 70 formats)
- Capture a screen
- Print an image
- Organize your graphics files

### Graphic Workshop

The Graphic Workshop is a shareware program (\$40) for MS Windows. It can be found at <http://www.muhimbi.com/Products/Graphic-Workshop.aspx>. It is really nice. It can convert a lot of graphic formats including .wmf.

### GIFConverter 2.3.7

Kevin A. Mitchell <[74017.2573@compuserve.com](mailto:74017.2573@compuserve.com)>

Kevin A. Mitchell P.O. Box 803066 Chicago, IL 60680-3066 USA

License: Shareware (\$40 + shipping)

GIFConverter, by Kevin A. Mitchell, reads and writes the following graphics file formats: GIF, MacPaint, PICT, RIFF, RLE, Thunderscan, Startup Screen, TIFF and JPEG (with or without QuickTime). In addition, it writes EPSF files. Also provided are image enhancement, cropping, color table selection, and dithering features.

### GraphicConverter 2.0.2

Thorsten Lemke <[thorsten\\_lemke@pe2.escape.de](mailto:thorsten_lemke@pe2.escape.de)>

Thorsten Lemke Insterburger Str. 631228 Peine Germany

License: Shareware (\$35)

GraphicConverter, by Thorsten Lemke, imports PICT, Startup-Screen, MacPaint, TIFF (uncompressed, packbits, CCITT3/4 and lzw), RIFF, PICS, 8BIM, 8BPS/PSD, JPEG/JFIF, GIF, PCX/SCR,

GEM-IMG/-XIMG, BMP (RLE compressed BMP & also), ICO/ICN, PIC (16 bit), FLI/FLC, TGA, MSP, PIC (PC Paint), SCX (ColorIX), SHP, WPG, PBM/PGM/PPM, CGM (only binary), SUN (uncompressed), RLE, XBM, PM, IFF/LBM, PAC, Degas, TINY, NeoChrome, PIC (ATARI), SPU/SPC, GEM-Metafile, Animated NeoChrome, magic, ImageLab/PrintTechnic, HP-GL/2, FITS, SGI, DL, XWD, WMF, Scitex-CT, DCX and KONTRON.

GraphicConverter exports PICT, Startup-Screen, MacPaint, TIFF (uncompressed, packbits and lzw), GIF, PCX, GEM-IMG/-XIMG, BMP, IFF/LBM, TGA, PSD, JPEG/JFIF, HP-GL/2, EPSF, Movie (QuickTime), SUN, PICS, PICT in Resource and PBM/PGM/PPM.

### JPEGView 3.3

Aaron Giles <[giles@med.cornell.edu](mailto:giles@med.cornell.edu)> Aaron Giles

182 E. 95th Street 11E New York, NY 10128 USA

License: Freeware (Send a Postcard)

JPEGView, by Aaron Giles, is a flexible image utility designed to allow quick, high-quality viewing of the most

common image formats, including JPEG, JFIF, GIF, PICT, Baseline and LZW-compressed TIFF, Windows BMP, StartupScreen, and MacPaint.

JPEGView can also convert between QuickTime JPEG and JFIF-standard JPEG files. Version 3.0 added a number of substantial new features, including full AppleScript support, new high-quality dithering routines, a greatly improved slide show, floating windows, etc.

JPEGView is now distributed only as a "fat binary", i.e. a program containing both 68K and Power PC code, thereby enabling it to run flat-out on both normal Macintoshes and the new Power Macintoshes.

Paint Shop Pro It is shareware. I found it on CIS

## Character Sets and Translation

### 1. Introduction

Text characters in an RTF file may be specified as literal characters or using notation `<i>xx</i>`, here `i>xx</i>` is the hex value of the character. RTF files also contain a control word that specifies the character set that's used within the document and governs the interpretation of character values. The charset control words are:

```
\ansi  ANSI (default) - Used by Word for Windows
\mac   Apple Macintosh
\pc    IBM PC
\pca   IBM PC page 850, used by IBM Personal System/2
```

Although the four charsets don't appear to differ for characters in the ASCII range (below 128), they differ considerably above the ASCII range (128-255). For example, the ANSI, Macintosh, and PC charsets represent the degree sign ("deg.") as `\b0`, `\a1`, and `\f8` respectively. Furthermore, even for a given charset, character values in the Symbol font represent different characters than they do generally. For example "a" in Symbol font is the greek letter alpha.

rtftohtml uses a translation model that divides translation into two parts. When the RTF file is read each character is mapped to an standard character name. This mapping is controlled by the files:

ansi-gen	Input mapping for ANSI, all fonts except symbol
ansi-sym	Input mapping for ANSI, symbol font
mac-gen	Input mapping for Macintosh, all fonts except symbol
mac-sym	Input mapping for Macintosh, symbol font
pc-gen	Input mapping for IBM PC, all fonts except symbol
pc-sym	Input mapping for IBM PC, symbol font
pca-gen	Input mapping for IBM PS2, all fonts except symbol
pca-sym	Input mapping for IBM PS2,, symbol font

General and symbol charset maps are stored in the text files `<i>ansi-gen</i>`, `<i>ansi-sym</i>`, `<i>mac-gen</i>`, `<i>mac-sym</i>`, `<i>pc-gen</i>`, `<i>pc-sym</i>`, `<i>pca-gen</i>`, and `<i>pca-sym</i>`. Each line of a charset file associates an RTF character value (field 2) with the standard character name to which the RTF character corresponds (field 1). Here's a sample from `<i>ansi-gen</i>`:

parenleft	(
parenright	)
space	" "
quotedbl	""
quoteright	"""
quoteleft	"`"

a	a
b	b
c	c
bullet	0x95
emdash	0x96
endash	0x97

Character values may be given as a single character (in which case the ASCII value is used), or as a hex number 0x<i>yy</i>. Single or double quotes may be used to quote values containing whitespace or quotes (e.g., use single quotes to quote a double-quote) Lines with a "#" in column one are taken as comments. Comments and blank lines are ignored.

When the HTML file is written, each standard character name is mapped to an ascii string. This mapping is controlled by the file html-map. The format of this file is field 1 is the standard character name, just as in the charset maps; field 2 is the output sequence to produce for the character named in field 1 Lines with a "#" in column one are taken as comments. Comments and blank lines are ignored.

Here's part of the file:

a	a
b	b
c	c
ampersand	&amp;amp;
less	&amp;lt;
equal	=
greater	&amp;gt;
trademark	(TM)
AE	&amp;AElig;
Aacute	&amp;Aacute;
Acircumflex	&amp;Acirc;

## Error Conditions and Remedies

First Things First - Determine the character set.

1. Identify which character set you are using by looking at the top of the RTF file. You should see one of the character sets identifiers, either \ansi, \mac, \pc, or \pca . If you do not see any of these then \ansi is your character set.

Error: Invalid Character Code \xx - outputting &amp;#000;

If you get this error, that means that a character appeared in the RTF input that has no standard character name. You fix this by adding that code (in this case the Hex Code xx) to your input character mapping file. Which file you use depends on your character set:

\ansi	ansi-gen and ansi-sym
\pc	pc-gen and pc-sym
\mac	mac-gen and mac-sym
\pca	pca-gen and pca-sym

Determine what character set you are using as described above find the character code in your source and determine what it looks like when it is displayed by your word processor. Also check to see if it is a character in the symbol font.

Add \e4 to the input translation file:

If you are using ansi and the character code is in the symbol font then you should edit ansi-sym and add:

```
somename      0xe4
```

The name that you use depends on what e4 looks like on your screen. If it is a bullet, then use

```
bullet        0xe4
```

This tells the filter to treat this character code as a bullet.

The complete list of standard character names is here:

Error: No output translation for :bullet

If you get this error, this means that the filter has no output mapping for the standard character name "bullet". To fix this, edit the html-map file and add a line describing what to output for a bullet. For example:

```
bullet *
```

What to do when you want to change a character translation

Start with html-map. Find the standard character name that you want to have translated, and then edit the corresponding string.

Standard Character Names

The complete list of standard character names is:

```
"nothing",  
"space",  
"exclam",  
"quotedbl",  
"numbersign",  
"dollar",  
"percent",  
"ampersand",  
"quoteright",  
"parenleft",  
"parenright",  
"asterisk",  
"plus",  
"comma",  
"hyphen",  
"period",  
"slash",  
"zero",  
"one",  
"two",  
"three",  
"four",  
"five",
```

"six",  
"seven",  
"eight",  
"nine",  
"colon",  
"semicolon",  
"less",  
"equal",  
"greater",  
"question",  
"at",  
"A",  
"B",  
"C",  
"D",  
"E",  
"F",  
"G",  
"H",  
"I",  
"J",  
"K",  
"L",  
"M",  
"N",  
"O",  
"P",  
"Q",  
"R",  
"S",  
"T",  
"U",  
"V",  
"W",  
"X",  
"Y",  
"Z",  
"bracketleft",  
"backslash",  
"bracketright",  
"asciicircum",  
"underscore",  
"quoteleft",  
"a",  
"b",  
"c",  
"d",  
"e",  
"f",  
"g",



"h",  
"i",  
"j",  
"k",  
"l",  
"m",  
"n",  
"o",  
"p",  
"q",  
"r",  
"s",  
"t",  
"u",  
"v",  
"w",  
"x",  
"y",  
"z",  
"braceleft",  
"bar",  
"braceright",  
"asciitilde",  
"exclamdown",  
"cent",  
"sterling",  
"fraction",  
"yen",  
"florin",  
"section",  
"currency",  
"quotedblleft",  
"guillemotleft",  
"guilsinglleft",  
"guilsinglright",  
"fi",  
"fl",  
"endash",  
"dagger",  
"daggerdbl",  
"periodcentered",  
"paragraph",  
"bullet",  
"quotesinglbase",  
"quotedblbase",  
"quotedblright",  
"guillemotright",  
"ellipsis",  
"perthousand",  
"questiondown",

"grave",  
"acute",  
"circumflex",  
"tilde",  
"macron",  
"breve",  
"dotaccent",  
"dieresis",  
"ring",  
"cedilla",  
"hungarumlaut",  
"ogonek",  
"caron",  
"emdash",  
"AE",  
"ordfeminine",  
"Lslash",  
"Oslash",  
"OE",  
"ordmasculine",  
"ae",  
"dotlessi",  
"lslash",  
"oslash",  
"oe",  
"germandbls",  
"Aacute",  
"Acircumflex",  
"Adieresis",  
"Agrave",  
"Aring",  
"Atilde",  
"Ccedilla",  
"Eacute",  
"Ecircumflex",  
"Edieresis",  
"Thorn",  
"Uacute",  
"Ucircumflex",  
"Udieresis",  
"Scaron",  
"Egrave",  
"Eth",  
"Iacute",  
"Icircumflex",  
"Idieresis",  
"Igrave",  
"Ntilde",  
"Oacute",  
"Ocircumflex",

"Odieresis",  
"Ograve",  
"Otilde",  
"Scaron",  
"Thorn",  
"Uacute",  
"Ucircumflex",  
"Udieresis",  
"Ugrave",  
"Yacute",  
"Ydieresis",  
"aacute",  
"acircumflex",  
"adieresis",  
"agrave",  
"aring",  
"atilde",  
"brokenbar",  
"cedilla",  
"copyright",  
"degree",  
"divide",  
"eacute",  
"ecircumflex",  
"edieresis",  
"egrave",  
"eth",  
"iacute",  
"icircumflex",  
"idieresis",  
"igrave",  
"logicalnot",  
"minus",  
"multiply",  
"ntilde",  
"oacute",  
"ocircumflex",  
"odieresis",  
"ograve",  
"onehalf",  
"onequarter",  
"onesuperior",  
"otilde",  
"plusminus",  
"registered",  
"thorn",  
"threequarters",  
"threesuperior",  
"trademark",  
"twosuperior",

"uacute",  
"ucircumflex",  
"udieresis",  
"ugrave",  
"yacute",  
"ydieresis",  
"Alpha",  
"Beta",  
"Chi",  
"Delta",  
"Epsilon",  
"Phi",  
"Gamma",  
"Eta",  
"Iota",  
"Kappa",  
"Lambda",  
"Mu",  
"Nu",  
"Omicron",  
"Pi",  
"Theta",  
"Rho",  
"Sigma",  
"Tau",  
"Upsilon",  
"varUpsilon",  
"Omega",  
"Xi",  
"Psi",  
"Zeta",  
"alpha",  
"beta",  
"chi",  
"delta",  
"epsilon",  
"phi",  
"varphi",  
"gamma",  
"eta",  
"iota",  
"kappa",  
"lambda",  
"mu",  
"nu",  
"omicron",  
"pi",  
"varpi",  
"theta",  
"vartheta",

"rho",  
"sigma",  
"varsigma",  
"tau",  
"upsilon",  
"omega",  
"xi",  
"psi",  
"zeta",  
"nobrksp",  
"nobrkhyphen",  
"lesseq",  
"greatereq",  
"infinity",  
"integral",  
"notequal",  
"radical",  
"radicalex",  
"approx",  
"apple",  
"partialdiff",  
"opthyphen",  
"formula",  
"lozenge",  
"universal",  
"existential",  
"suchthat",  
"congruent",  
"therefore",  
"perpendicular",  
"minute",  
"club",  
"diamond",  
"heart",  
"spade",  
"arrowboth",  
"arrowleft",  
"arrowup",  
"arrowright",  
"arrowdown",  
"second",  
"proportional",  
"equivalence",  
"arrowvertex",  
"arrowhorizex",  
"carriagereturn",  
"aleph",  
"Ifraktur",  
"Rfraktur",  
"weierstrass",

"circlemultiply",  
"circleplus",  
"emptyset",  
"intersection",  
"union",  
"propersuperset",  
"reflexsuperset",  
"notsubset",  
"propersubset",  
"reflexsubset",  
"element",  
"notelement",  
"angle",  
"gradient",  
"product",  
"logicaland",  
"logicalor",  
"arrowdblboth",  
"arrowdblleft",  
"arrowdblup",  
"arrowdblright",  
"arrowdbldown",  
"angleleft",  
"registersans",  
"copyrightsans",  
"trademarksans",  
"angleright",  
"mathplus",  
"mathminus",  
"mathasterisk",  
"mathnumbersign",  
"dotmath",  
"mathequal",  
"mathtilde",  
"newline",

This document was created from RTF source by rtfhtml version 2.7.  
rtfhtml Change History

rtf-ctrl has been Revised

When Word 6.0 was released, it introduced some new control words which rtfhtml 2.7.5 was not aware of. These caused messages to appear like:

ReadStyleSheet: unknown token "\widctlpar"

Unknown symbol "\sprstsm" near line 9, position 66.

Unknown symbol "\truncex" near line 9, position 74.

Unknown symbol "\nolead" near line 9, position 81.

Unknown symbol "\msmcap" near line 9, position 88.

Unknown symbol "\widctlpar" near line 12, position 17.

In addition, Word 6.0 uses \super for footnotes instead of \up, which confuses the filter. To correct these problems, get the new version of rtf-ctrl and copy it into your folder or directory containing the filter.

#### Changes in rtfhtml 2.7.5

Changed the processing of line breaks (\line in RTF, &lt;shift&gt;&lt;return&gt; in MS Word). These were being treated the same as paragraph marks, but they should have been treated as a simple break. Where

this makes a difference is in lists.

The filter now correctly renders the following example as a single list element (i.e. one bullet mark) across two lines.

- \* some list element\line  
more stuff belonging to same list element, but on a different line.

MSWord 6.0 for the Macintosh is now using 'RTF' as the file type for rtf files, instead of TEXT. This caused rtfhtml to ignore those files in the OPEN file dialog box. You could not drag and drop those RTF files onto the filter either. The filter now accepts file types of RTF or TEXT.

#### IMPORTANT

To make drag-n-drop of MSWord 6.0 RTF files work on your Macintosh, you must

1. Download rtfhtml 2.7.5 onto your Macintosh
2. Copy ALL PREVIOUS VERSIONS of rtfhtml onto a backup disk. Then REMOVE them from your hard-disk.
3. Restart your Macintosh, while holding down the OPTION and Apple keys. This will re-build your desktop.

#### Changes in rtfhtml 2.7.4<ul>

Fixed a bug where column widths of 1 character caused an infinite loop.

Fixed a problem for <b>NeXT</b> <b>RTF</b> files where \ followed by a newline was not being interpreted as a \par mark.

Footnote numbers were not being reset at the start of each file, so the second document converted would have footnotes starting at &gt; 1.

Some Word 6.0 RTF symbols were improperly coded as destinations. This caused rtfhtml to discard the rest of the file (searching for a closing '}'

#### Changes in rtfhtml 2.7.3

Modified status bar for <b>Mac</b> platform - it was incorrect for large files.

Changed error messages to print line and column information rather than context.

Modified <b>MWF</b> processing to produce the correct <b>WMF</b> header

Added code to recognize that style names of the form "Normal, fred" actually mean that either Normal or fred are recognized as the style name. (Implies that style names with ',' in html-trans are incorrect.)

Added -V option to UNIX version

Corrected usage message on UNIX version

#### Changes in rtfhtml 2.7

Footnotes are now written to a separate file (filename\_fn.html). This allows the &lt;back> button of all browsers to navigate back from a footnote to the footnote reference.

A table of contents file is now generated using the text from all headings &lt;H1>, ...&lt;Hn>, and any table of contents entries in the RTF. The table of contents has hypertext links from the TOC entries to their locations in the main HTML text. The -T option disables this feature.

Generation of graphics files can be disabled through the -G option. The filter still generates the hypertext links to the graphics files, but it does not re-write the files themselves. This improves performance for graphics intensive RTF files.

Fixed a bug where table of contents entries and index entries were not appearing in the HTML output.

Fixed a bug that caused aborts when graphics occurred in a table.

rtftohtml now closes off any "text" markup (&lt;em>,&lt;b>,&lt;cite>) before emitting a paragraph tag. This more closely matches the HTML DTD and corrects a few error conditions.

The format of the html-trans file has changed. You will need to use the html-trans file from this distribution and then add any changes that you made to customize your html-trans.

Footnotes no longer require special styles in order to work. This means that all superscripts can now be translated to &lt;sup>,&lt;/sup> or any other markup you prefer.

The Macintosh interface now allows the file Creator to be selected for graphics, HTML and error files.

Fixed a bug that caused formatted text in tables to be lost or added to the wrong cells.

## Notes on rtftohtml 2.5

This release of rtftohtml requires a new version of the html-trans file. Using your old one will not work.

The new format file has an additional item added to the table. An entry now looks like this:

```
#"name","starttag","endtag","col2mark","tabmark","parmark",allowtext,cannest,DelteCol1,fold.PTag
```

```
"Normal","", "\n", "\t", "\t", "&lt;p>\n", 1, 0, 0, 1,
"pre", "&lt;pre>", "&lt;/pre>", "\t", "\t", "\n", 0, 0, 0, 0
```

The field should be set to zero (0) for all entries except <tt>"pre"</tt> or <tt>"listing"</tt>. A zero entry allows rtftohtml to insert newlines <tt>"\n"</tt> into the source so that lines do not get too long in your HTML file. This feature is designed to make editing/viewing of HTML source files easier with editors like <i>vi</i>. You must use one (1) for "pre" and "listing" because newlines are significant characters for these markups. You must add the following entry to your <i>.PTag</i> table. This is required to support table translation.

```
# This is a required entry; tables will be formatted with this entry
"_Table", "&lt;pre>", "&lt;/pre>", "\t", "\t", "\n", 0, 0, 0, 0
```

You must add the following entry to your <i>.PMatch</i> table

```
# This is a required entry; tables will be formatted with this entry
"_Table", 0, "_Table"
```



## Table Translation

rtftohtml will now translate tables into preformatted ascii tables.

Tables cannot contain any markup. That is, **BOLD ITALIC**etc. will be ignored. Also tables cannot contain anchors, footnotes, or hidden text. The filter will produce a table that has the same number of columns as the input. The size of the columns will be proportional to the sizes of the input columns. Note that there is no guarantee that text that fit in your original table will still fit in the formatted version. This is because the point size of input may not be proportional to the fixed size 80-characters per-column output. If you find that you do not like the look of the formatted table - you must re-size the columns in your RTF source. Columns are always separated by two blanks. Columns can be left justified, right justified or centered. Merged cells are allowed. There is `<b>no</b>` support for decimal aligned data. This just looked too hard and prone to error. For now, you will have to make sure that all your data has the same number of trailing digits and stick to right justified text.

**A REAL Macintosh interface** The filter now supports drag-n-drop on the Macintosh!  
Options are set with menu selections! The signature is even registered with Apple!

To use rtftohtml on the Macintosh: Keep all of the translation files (html-trans, ansi-map...) in the same folder as the application.

Drag any RTF file onto rtftohtml (requires system 7.)

The output files will be placed in the same folder as your input file. **</b>** If there are any errors in

translating your file, the error messages will be put into a file called `ilename.err`. Some errors will cause rtftohtml to exit, others will not. Once rtftohtml is started, it will continue to run **</b>**! This allows you to

translate many documents at a time. You can drag lots of files onto rtftohtml and it will translate all of them (unless it encounters a fatal translation error.) To use non-default translation options, simply start rtftohtml by double clicking on it. Then choose your option from the options menu. These settings will remain in effect until rtftohtml exits. **<b>**The next time you run rtftohtml you must choose your option settings again. You can open RTF files using **<b>**Open**</b>**... from the **<b>**File**</b>** menu.

This will only work for files with an extension of .rtf (or .RTF). If you want to open a file without an extension of .rtf, you must hold down the option key when selecting Options Menu. The Graphic file extension in the Options menu allows you to change the default extension for graphic files from ".gif" to an extension of your own choosing. The Inline graphics option will (when set) use `&lt;IMG SRC=...` markup in HTML files for graphics instead of `&lt;A HREF=`.

The Auto Filenames option when set will put the HTML output in the same folder as the input, with a .HTML extension. If you want to specify your own filenames, turn this option off. Then rtftohtml will prompt you for an output file name each time a file is translated.

## Miscellaneous Features

All paragraph styles that appear in a document but that do not appear in the html-trans file will be treated as a warning. This provides you with a list of paragraph styles that you can then add to html-trans.

You can tell rtftohtml to discard all of the text with a given paragraph style. This is done by using the "\_Discard" in the .PMatch table. For example the following entries will discard all text with paragraph styles "toc 1-5" and "index 1-5". (Word uses these paragraph styles for

generating a table of contents and index. Since these contain page numbers which do not apply to HTML documents, you may wish to discard them.)

```
"toc 1",0,"_Discard"  
"toc 2",0,"_Discard"  
"toc 3",0,"_Discard"  
"toc 4",0,"_Discard"  
"toc 5",0,"_Discard"  
"index 1",0,"_Discard"  
"index 2",0,"_Discard"  
"index 3",0,"_Discard"  
"index 4",0,"_Discard"  
"index 5",0,"_Discard"
```

Bugs Fixed in 2.5rtftohtml was generating "unknown symbol" errors for optional destinations (RTF with a markup of {\*\xxx ... } The convention for RTF readers is that any optional destination that a reader does not understand should be silently

#### Bugs fixed in 2.1

rtftohtml generates empty anchors <a name="myname"></a>, where the "Named Anchor" feature is used. Xmosaic does not support empty anchors, although other browsers (including Mac Mosaic) do. I have reported this as a bug to the good folks at NCSA. I am hoping that they fix this so that I don't have to patch the filter. Stay tuned. - The good folks at NCSA told me not to wait, so I revised rtftohtml to create anchors with an extra space:

<a&nbsp;name="myname"> </a>

Imbedded object were not appearing in the HTML output. Now they are!

Fixed the usage line to reflect that -H is not a legal option.

Windows Metafiles were given an extension of ".meta". That has been changed to ".wmf".

The Mac version failed to find html-trans, html-map when the input file was not in the same folder.

Nested lists were not working properly. How did I miss that one

#### Bugs fixed in 2.0.1

RTF files created without a leading /pard caused aborts Graphics within a HOT text area caused failures instead of generating IMG links Anchor names and Hot Text links were getting an extra ">" inserted.

Footnotes incorrectly got a "#" at the anchor points.

Footnotes reference marks were generated without any separating text or space.

Should be generated as [1] filter did not understand \tql which is generated in FrameMaker RTF output. This has been added to reader.c and rtf.h in the source directory.

Eventually, these changes will be incorporated in an official distribution of the RTF package.

(Binary versions of rtftohtml contain the fix) FrameMaker RTF assumes that \pard implies \plain - MS Word for the Mac

appears to handle it that way, so I have updated the filter.

### 3.3 MEGADOTS :

What's New, What's Hot -- August 13, 1995

Raised Dot Computing is preparing a mass-mailing of MegaDots updates. Raised Dot Computing is a few days away from mastering version 1.5a of MegaDots. As soon as possible, we will be mailing out version 1.5a updates. We will be mailing out free updates to the following classes of customers:

- MegaDots that were shipped from Raised Dot after May 4, 1995
- All those who have a copy of the MegaDots Network Key
- All those who have (or have had) Premium Support
- MegaDots beta testers
- Dealer Demonstration Copies

Purchasers of older MegaDots will receive letters asking if they want to purchase a MegaDots 1.5a update for \$50.

Built into MegaDots 1.5a is a program to install a MegaDots icon in Windows 3 or in Windows 95.

This whole Web page is new. Its existence is based on a new feature in MegaDots 1.5: the ability to import and export HTML files including cross-references. Inside MegaDots, all the links in the HTML documents can be executed. If you loaded 20 HTML files from the Web that all referenced each other, you could execute the links inside MegaDots. MegaDots is now an ideal off-line reader for Web documents for blind and sighted users. For more information, see Using MegaDots to Produce Web Documents.

At this point, all copies of MegaDots 1.5 sold come with the Spell Checker at no additional charge. Ordinarily, this is a \$35 additional item. We are also selling Duxbury trade-ins (including our Spell Checker of course) for \$175. These promotions expire at the end of 1995.

The biggest promotion we have is to sell all MegaDots with the Spell Checker at no additional charge.

MegaDots purchasers may want to check out the 1995 MegaDots Catalog. The catalog includes quite a few items that were never listed in previous newsletters or mailings (such as Marcia Leibowitz's transcriber manual, and the new Beginner Braille module).

Windows users may want to check out Using MegaDots from Windows 3 and Windows 95.

As many of you may know, Raised Dot Computing is working on an elaborate project to translate mathematical and technical text into braille mathematics (Nemeth braille code) and into verbal renditions through voice.

synthesis. For an up to date report on this project see Nemeth Code Project.

Overseas readers will be pleased to know that MegaDots now supports British braille in MegaDots 1.5. We also support the system used in Australia and South Africa (British braille up with capitalization marks). If you would like more information about the availability of British style or Australian style MegaDots, please contact Raised Dot Computing.

We welcome your comments and suggestions to further improve this Web document.

### Using to MegaDots to Prepare Web Documents

Using to MegaDots to Prepare Web Document MegaDots is the braille translator and word processor

from Raised Dot Computing. MegaDots version 1.5 now reads and writes HTML files, the kind used on files in the World Wide Web.

MegaDots imports about 100 different file formats, so ou can import ASCII files, Macintosh files, Windows files, WordPerfect files, or other HTML files.

Inside the MegaDots Editor, you can apply the appropriate styles on your material. The style manual (see below) tells you the MegaDots format which produces appropriate material on the World Wide Web.

When you are finished writing in MegaDots, use the control-F4 command to export the file. Choose HTML as the file format.

Now upload the HTML files to your Web directory somewhere on the Internet. Instant Web page.

### Style Manual

To create HTML files, you need MegaDots 1.5 beta, dated after April 26, 1995. If you need this update to read or write HTML files, contact Raised Dot Computing at 608-257-8833

Each MegaDots file prepared for the Web should start with a paragraph in the Header style (Alt-R H &lt;Enter&gt;), followed by a paragraph that is Heading level 1 (Alt-H &lt;Enter&gt;). The material in the MegaDots Header style becomes the HTML &lt;Title&gt;, the material in MegaDots Heading level 1 becomes HTML  
Avoid using emphasis in headings.  
Avoid using list style for really long text.

To make a horizontal line use an article terminator after a Body text paragraph.

o leave a bookmark in a file (for use with hypertext links). First write the text and then highlight it (MegaDots lets you highlight with the mouse or with control-X). Then type Alt-J followed by Insert. Type in the name of the bookmark. Be aware that you can create MegaDots bookmarks WITHOUT highlighting text, but only the bookmarks created WITH highlighting work as HTML bookmarks.

To make hypertext links, first write the text and then highlight it. Then type control-F J. (The control-F J is totally undocumented, and not listed in any menu.) You are given an opportunity to type the file name and the bookmark you want. If you want to start at the beginning of the file, you can leave the bookmark field blank. The file name should reflect how you want it to be on the Web. For example, if you want to reference the file BIGlist.html then put that in the file name field.

MegaDots can execute the hypertext links from within MegaDots. This lets you test out your local links. If you want to test out a link in MegaDots, it looks for a file with the same file name (but with the MEG file extension) in memory or in your current directory.

### Grabbing other People's Web Material

How can you obtain HTML files from the World Wide Web?

This is fairly easy. Send an e-mail message to

The message should consist of the linesource `http://URL`(you can have up to 10 lines of request per e-mail message). You may get your answer in a few minutes, or in a day. You can obtain any text you want from anywhere in the World Wide Web this way.

Once you get the e-mail message, strip off the header information, and import the file into MegaDots.

### Putting it all on the Well

The Well is an Internet provider used by Raised Dot Computing. They have a self-publishing system that is simple and inexpensive. Obviously, you can upload your HTML files onto one of hundreds of Internet providers. I am giving the data on the Well here to show how simple all this really is.

To dial up the Well, call 415-332-8410. You can connect through Telnet, but dial direct when you need to upload.

To create a web directory, type `!makeweb` at the Well prompt. This creates a directory called WEB (case is important).

Use the `!up` program to upload your files.

Upload your home page into `WEB/index.htm`. I use Procomm for communications. For uploads, I use X-modem protocol for textfiles (as opposed to binary files). Crude but effective.

Once you have all your HTML files on the Well, you want the rest of the world to be able to read them without having the right to modify them. There is a simple command `!websecurity` (use at the OK: Well prompt) that does this for all the files in your WEB directory.

Your URL is `www.well.com/www/user-id`

Since my user-id on the Well is `dnavy`, my URL is `www.well.com/www/dnavy`. That is all there is to it. The only other thing to do is to send announcements to anyone who is interested that you are now on the Web.

On different on-line systems, there are different commands to load material onto the World Wide Web. Usually, these are fairly simple and easy to learn. Good luck.

## Using to MegaDots to Prepare Web Documents

MegaDots is the braille translator and word processor from Raised Dot Computing. MegaDots version 1.5 now reads and writes HTML files, the kind used on files in the World Wide Web.

MegaDots imports about 100 different file formats, so you can import ASCII files, Macintosh files, Windows files, WordPerfect files, or other HTML files.

Inside the MegaDots Editor, you can apply the appropriate styles on your material. The style manual (see below) tells you the MegaDots format which produces appropriate material on the World Wide Web.

When you are finished writing in MegaDots, use the control-F4 command to export the file. Choose HTML as the file format.

Now upload the HTML files to your Web directory somewhere on the Internet. Instant Web page.

### Style Manual

To create HTML files, you need MegaDots 1.5 beta, dated after April 26, 1995. If you need this update to read or write HTML files, contact Raised Dot Computing at 608-257-8833

Each MegaDots file prepared for the Web should start with a paragraph in the Header style (Alt-R H it;Enter&gt;), followed by a paragraph that is Heading level 1 (Alt-H &lt;Enter&gt;). The material in the MegaDots Header style becomes the HTML &lt;Title&gt;, the material in MegaDots Heading level 1 becomes HTML

Avoid using emphasis in headings.

Avoid using list style for really long text.

To make a horizontal line use an article terminator after a Body text paragraph.

To leave a bookmark in a file (for use with hypertext links). First write the text and then highlight it (MegaDots lets you highlight with the mouse or with control-X). Then type Alt-J followed by Insert. Type in the name of the bookmark. Be aware that you can create MegaDots bookmarks WITHOUT highlighting text, but only the bookmarks created WITH highlighting work as HTML bookmarks.

To make hypertext links, first write the text and then highlight it. Then type control-F J. (The control-F J is totally undocumented, and not listed in any menu.) You are given an opportunity to type the file name and the bookmark you want. If you want to start at the beginning of the file, you can leave the bookmark field blank. The file name should reflect how you want it to be on the Web. For example, if you want to reference the file BIGlist.html then put that in the file name field.

MegaDots can execute the hypertext links from within MegaDots. This lets you test out your local links. If you want to test out a link in MegaDots, it looks for a file with the same file name (but with the MEG file extension) in memory or in your current directory.

## Grabbing other People's Web Material

How can you obtain HTML files from the World Wide Web?

This is fairly easy. Send an e-mail message to

agora@mail.w3.org<. The message should consist of the line: source http://URL (you can have up to 10 lines of request per e-mail message). You may get your answer in a few minutes, or in a day.

You can obtain any text you want from anywhere in the World Wide Web this way.

Once you get the e-mail message, strip off the header information, and import the file into MegaDots.

Putting it all on the Well

The Well is an Internet provider used by Raised Dot Computing. They have a self-publishing system that is simple and inexpensive. Obviously, you can upload your HTML files onto one of hundreds of Internet providers. I am giving the data on the Well here to show how simple all this really is.

To dial up the Well, call 415-332-8410. You can connect through Telnet, but dial direct when you need to upload.

To create a web directory, type!makewebat theOKWell prompt. This creates a directory called WEB

(case is important).

Use the !up program to upload your files.

Upload your home page intoWEB/index.html I use Procomm for communications. For uploads, I use X-modem protocol for textfiles (as opposed to binary files). Crude but effective.

Once you have all your HTML files on the Well, you want the rest of the world to be able to read them without having the right to modify them. There is a simple command !websecurity(use at theOK: Well prompt) that does this for all the files in your WEB directory.

On different on-line systems, there are different commands to load material onto the World Wide Web. Usually, these are fairly simple and easy to learn. Good luck.

### 3.4. DOTSPLUS

#### Introduction to Dotsplus

Dotsplus is a combination of 8-dot braille <EM> plus</EM> raised representations of some print symbols. Braille cells are used for the usual characters, numbers, and punctuation, but raised images are used

for fraction bars and integral signs and other symbols. The print symbols used are normally enlarged by a factor of 2.5-3 and then printed out as raised images so that they are the same height as the braille dots.

This makes possible the use of a wider range of symbols than allowed by the 256 possible 8-dot braille cells but more importantly the use of raised images enhances the ability to lay out the page similar to print

pages. For example a fraction can be represented in Dotsplus as two braille characters, one above the other, with a raised fraction bar between them. Since none of the World Wide Web clients are yet able to display raised images, an example of Dotsplus can be obtained with the help of the U.S. Postal Service by writing to

gardner@zircon.physics.orst.edu. For the sighted, a bitmap image showing an example of Dotsplus is displayed below, <IMG SRC= "dpeqr.gif" ALT = "bitmap image of Dotsplus"> along with an image of the regular print version of the same equation <IMG SRC= "dpeqc.gif" ALT = "bitmap image of Dotsplus">.

A few papers have been published about Dotsplus and they are also available here.

csun93.html">Dotsplus-Better than Braille?

"csun94.html">Accessibility to Scientific Information by the Blind: Dotsplus and Aster could make it asy

John Gardner,

gardner@zircon.physics.orst.edu

Bill Barry,

barryw@ucs.orst.edu

#### Dotsplus - Better than Braille?

John A. Gardner

Department of Physics

Oregon State University

Corvallis, OR 97331

Dotsplus is an experimental method of printing technical literature for blind readers. It incorporates both braille and graphics in an integrated fashion. At present, dotsplus documents can be produced with commercial presses or capsule paper. Desk top printing is possible in principle with a wax-jet printer or high-resolution braille embosser, but no fully satisfactory models are currently being marketed. If useful printers are introduced, raightforward font sets and software drivers should allow dotsplus documents to be produced with common word processors or from such programs as TeX or SGML, both of which are accessible to blind users.



Examples in this document were produced by a sighted person using Microsoft Word with special dotsplus fonts.

Math equations are reproduced as enlarged images of the equation as printed for a sighted person. The plus, minus, and equals signs, parentheses, brackets, the division line in fractions, and the multitude of exotic symbols used in advanced science and math are all printed as raised images. Some symbols may be enlarged more than others or their shapes may be emphasized for clarity, but all are instantly recognizable by sighted mathematicians. Braille is used only for alphabetic characters, numbers, and punctuation marks. The placement of symbols and characters is the same in dotsplus documents as it would be for a sighted person - subscripts are dropped below and superscripts are raised above the position of the main character; numerators of fractions are placed above the denominators; in advanced equations, symbol placement such as sum and integral limits, are preserved.

Text in a dotsplus document can be written in standard braille or dotsplus symbols. Dotsplus text may be preferable if the text contains many scientific symbols.

In order to avoid ambiguity and to incorporate most of the characters used in elementary math into single braille cells, dotsplus braille is based on the 8-dot cell and does not use any dropped symbols. Dotsplus braille cells have been chosen to be as intuitively-related as possible to literary braille. The lower case alphabet is standard braille. Capitalization is indicated by an additional dot on the lower left (usually called the dot-7 position). Greek characters, which are often used in algebra, are indicated by an additional dot in the lower right (usually called the dot-8 position). For example, a capital l is indicated by a dot 1,2,3,7 symbol. A lambda is indicated by a dot 1,2,3,8 symbol, and a capital lambda by a dot 1,2,3,7,8 symbol.

Dotsplus adopts the European computer braille standard to indicate numbers as single braille cells. In this standard, the digits 1-9 are given by the braille cells for the letters a-i with an added dot 6. For example, 1 is the cell dot 1,6. Seven is the cell dot 1,2,4,5,6. The digit 0 cannot be represented by a j with a dot 6, since this is the letter w. Instead 0 is represented by a dot 3,4,6 cell.

Dotsplus punctuation marks are a combination of graphics and braille. The usual literary 4-dot dropped cells are used with a graphics indicator directly above them. Several possibilities for the indicator symbol are still being tested. For example, a comma, usually represented by a dot 2 cell is represented in dotsplus by a square with a dot just below the left bottom corner of the square. For maximum international consistency, the original Louis Braille dropped cell punctuation definitions are used. Only the question mark is different from standard English language braille

Not all mathematics can be represented with single braille cell characters. Bold, script, gothic, hebrew, etc. characters require two cells. These pose no particular difficulty but do increase the size of equations that contain large numbers of such symbols

The single-cell dotsplus braille

set is illustrated in Figure 1. The dotsplus image of a page from a high school algebra book is shown in Figure 2. The same image but overprinted with the ink print equivalent is shown in Figure 3. The ink print page and dot image page are obtained from the same word processor file by a font change. A sighted computer user with no knowledge of braille can prepare such pages.

Ideally, all dotsplus documents would have raised images overprinted by the ink print equivalent as illustrated in Figure 3.

This technique allows both visual and tactile reading by visually-impaired people and enhances communication between blind students and their sighted peers and teachers. The author believes that dotsplus could become the preferred method for printing advanced mathematical and scientific literature for blind people. Portions of several high school and college textbooks are presently being prepared in dotsplus for testing by students who are using those books in their classes. The design of dotsplus has been developed after extensive testing and criticism by blind scientists, educators, and braille authorities in the US and several foreign countries. Blind scientists had little difficulty learning the capitalization and greek indicator convention but believe that some practice will be required for users to become familiar with dotsplus numbers and use of subscripts, superscripts, etc. \* Research program supported in part by the US National Science Foundation

Dotsplus braille cell conventions.

Figure 2 Dotsplus raised image page transcribed from Advanced Algebra (University of Chicago School Mathematics Project), Scott Foresman Publishers, Glenview, Illinois.

Figure 3. Raised image from Figure 2 in light gray with equivalent ink print in dark gray. Ideal dotsplus literature would have inkprint as well as raised images, so it can be read easily by sighted and blind readers.



# ***votre interlocuteur***

---

## **E**xpérience professionnelle

EVER met à votre disposition une équipe d'ingénieurs, d'informaticiens et de formateurs hautement qualifiés.

Grâce à leur savoir-faire acquis par une expérience professionnelle de plus de 10 ans, ce sont de véritables spécialistes du traitement automatisé de l'information.

## **V**ision globale

EVER vous propose une palette complète de services depuis l'étude de faisabilité, l'élaboration du cahier des charges jusqu'à la formation des utilisateurs en passant par le développement logiciel, sa mise en exploitation et le service après-vente.

## **E**coute personnalisée

La volonté première d'EVER sera toujours :

- . L'écoute de vos besoins pour mettre à votre disposition des outils opérationnels et fiables ;
- . la prise en compte de votre environnement informatique pour combiner au mieux, innovations technologiques et objectifs économiques de rentabilité.

## **R**éponse adaptée

EVER vous apporte une réponse adaptée et des solutions cohérentes et concrètes à tous les aspects spécifiques de votre chaîne de traitement de l'information.

# *Gestion Electronique de Documents*

---

## **R**apidité de mise en œuvre

DORIS est un logiciel convivial et simple d'utilisation. Le multifenêtrage et la souris permettent de maîtriser très facilement le fonctionnement du logiciel.

Des écrans d'aide sont disponibles à tout moment.

Les rubriques sont multivaluées et en format variable.

La création d'index sur une ou plusieurs rubriques, avec ou sans liste de mots vides, est un gage de fiabilité et de rapidité de la recherche.

DORIS propose un mode de recherche en langage simple. Opérateurs booléens et de comparaison, troncatures, historiques d'interrogation, recherche par étapes, profils, thésaurus, enrichissent les possibilités de recherche.

## **I**ntégration de trois systèmes

DORIS allie la puissance du relationnel, la finesse du documentaire et la souplesse de programmation de la conception objet.

En intégrant ces 3 visions, DORIS est un véritable outil d'aide à la description de systèmes complexes, un logiciel qui fait le lien entre les sciences de l'information et les méthodes d'organisation. DORIS organise l'information.

La structuration hiérarchique des données et les liens multi-fichiers font de DORIS un système de représentation statique et dynamique de la réalité qui nous entoure. Ce système est rendu vivant par des techniques bien plus intuitives que les solutions informatiques traditionnelles.

## **S**écurité des données

DORIS garantit la confidentialité des données stratégiques et contrôle l'intégrité des données. L'accès aux informations présente 3 niveaux de sécurité: UNIX, SGBDR et DORIS.

DORIS prévoit des procédures de recouvrement en cas d'incident technique.

La pérennité du système est assurée par le choix des Normes Internationales (ISO) et des grands standards du marché :

- UNIX
- SQL
- ASN1
- X-WINDOW
- ETHERNET.

*Système innovant de Gestion Electronique de Documents, DORIS est aussi un puissant logiciel documentaire.*

*DORIS s'adapte à vos besoins spécifiques sans bouleverser l'organisation existante.*

## Diversité des applications

Entièrement paramétrable, DORIS permet de définir simplement des applications variées quel que soit le volume de données à traiter :

- archives
- bibliothèques
- brevets
- cahiers des charges
- catalogues produits
- circulaires
- contrats
- courrier
- déclarations
- documentation technique
- dossiers
- médiathèque
- polices d'assurances
- revue de presse

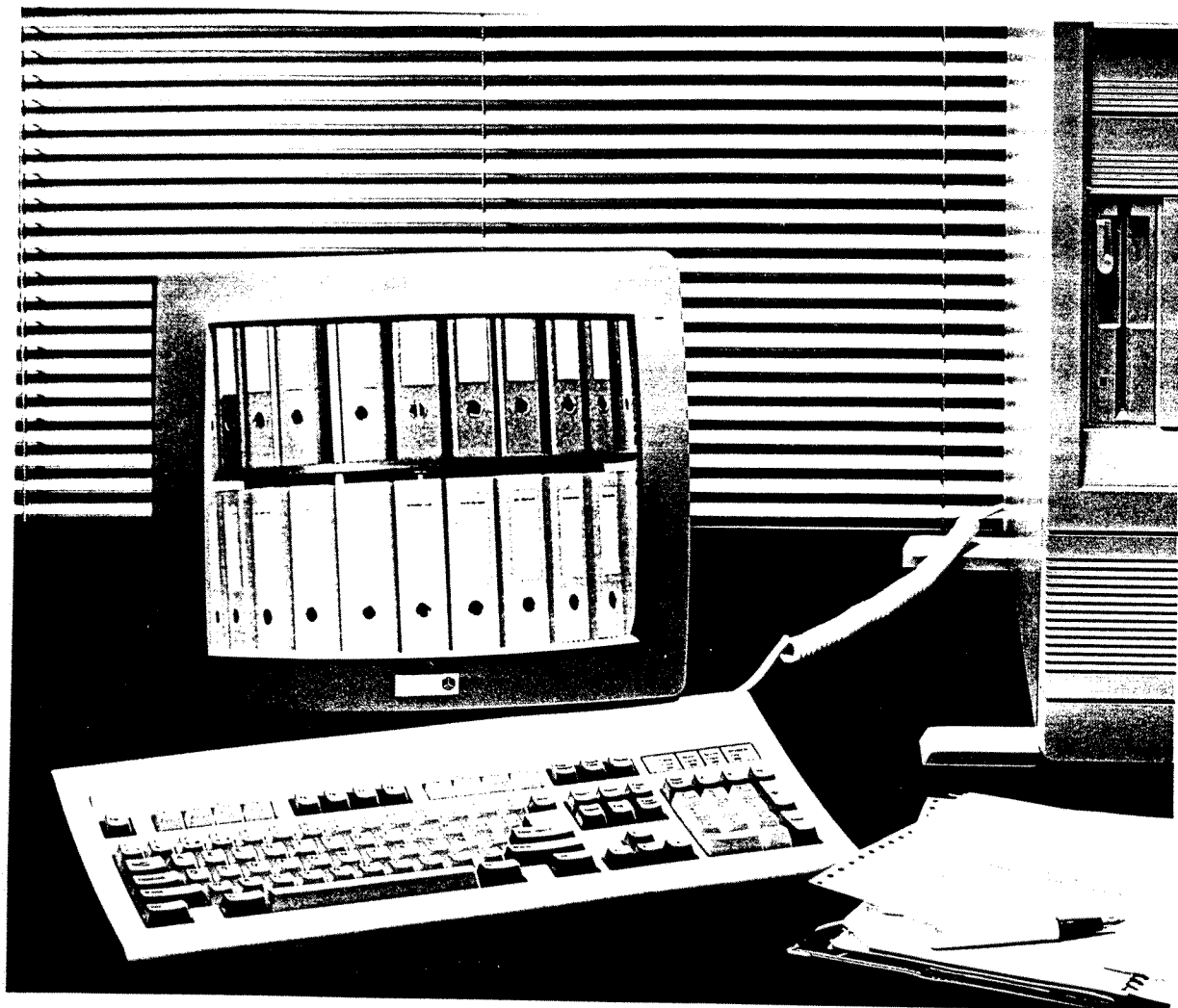
et plus généralement, toute application de Gestion Electronique de Documents.

## Ouverture vers l'extérieur

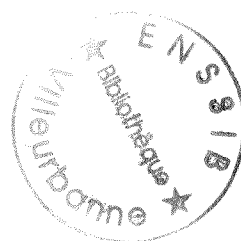
DORIS offre la possibilité d'importer des fichiers externes (traitement de texte, tableurs, logiciels documentaires, base de données etc.).

A l'inverse, DORIS sait exporter des données vers d'autres logiciels (CAO, PAO, SGBDR). Ces fonctions d'échange en font un outil ouvert et performant.

DORIS intègre les nouvelles technologies d'archivage. Fonctionnant avec des interfaces scanner et/ou Disque Optique Numérique de type WORM, DORIS permet l'archivage de données numérisées multimédia : textes, images et sons. Il permet la compression des données aux normes G3 et G4.



**Doris<sup>®</sup>**  
*from*  
**EVER**



*La valorisation*  

---

*du*  

---

*Capital Information*

## INTRODUCTION

### PRESENTATION DE LA SOCIETE EVER

Créée en 1990 par une équipe de spécialistes du traitement automatisé de l'information, la société EVER est une SSII (Société de Services en Ingénierie Informatique) spécialisée dans la mise en place de programmes informatiques de Gestion Documentaire et de Gestion Electronique de Documents.

Basée à Lyon au 76 rue de Bellecombe dans le 6<sup>e</sup> arrondissement, lieu de son siège social, EVER dispose d'un bureau à Paris. La société a mis en place dès sa création, des accords de coopération et de partenariat internationaux pour la diffusion de ses produits sur le marché international. Bénéficiant d'une expérience de son personnel de plus de quinze années dans le secteur de l'informatique documentaire, la société EVER conçoit des logiciels exclusivement dans ce créneau, et commercialise ses solutions.

Société de Services, indépendante de toute source ou fournisseur privilégié, la société EVER conseille et assiste librement ses clients lors des différentes étapes qui conduisent à la mise en place d'un Système de Gestion Documentaire ou de Gestion Electronique de Documents depuis l'étude de sa faisabilité, l'élaboration du cahier des charges, en passant par l'adaptation à ses besoins jusqu'à la formation des utilisateurs, la mise en exploitation et le service après-vente personnalisé.

La société EVER a développé un logiciel de Gestion Electronique de Documents et de Gestion Documentaire (DORIS) et s'est lancée dans un développement spécifique de Gestion de Bibliothèque (LORIS).

## **PRESENTATION DU LOGICIEL DORIS**

DORIS (Documentary Object-oriented Relational Information System) est une synthèse intégrée des méthodes d'organisation et des sciences de l'information. Ce logiciel permet de décrire simplement une réalité complexe et s'adapte tout à fait au stockage de grandes masses d'information, de nature très variée (document numérique, document éditeur ou traitement de textes, fax, télex, image analogique image vidéo, document sonore, document graphique).

Il est entièrement paramétrable, pour ce qui est des structures de fichiers et structures de fiches. Il est donc possible de définir exactement le modèle de fiche signalétique le plus approprié.

DORIS comporte deux chaînes de traitement parallèles :

### **La chaîne Gestion Documentaire**

- . Gestion de la base de données documentaire
- . Consultation de la base de données documentaire
- . Visualisation et impression des fiches signalétiques
- . Sécurité des données

### **La chaîne Gestion Electronique de Documents et Archivage .**

- . Scannérisation ou Numérisation des documents
- . Stockage des documents sur serveur d'archivage
- . Consultation de la base de documents archivés
- . Visualisation et impression des documents archivés
- . Sécurité des documents archivés



BIBLIOTHEQUE DE L'ENSSIB



966551G