



# Rapport de stage

IUP MIAGE 2ème Année



**Julie Pham**  
**Septembre 1999**

*(Stage effectué avec le soutien financier du Département SPM du CNRS  
(attribution de 2 mois d'allocations de stage)*

## Table des matières

Introduction	3
<i>I. Contexte dans lequel le travail a été effectué</i>	4
I.1. Présentation de la Cellule MathDoc	4
I.2. Le Catalogue Fusionné des périodiques de mathématiques	5
I.3. Une gestion complexe	7
<i>II. Sujet proposé</i>	8
<i>III. Travail réalisé</i>	9
III.1 Cahier des charges	9
III.2 Réalisation	13
III.3 Implémentation du CGI	16
<i>IV. Intérêt du stage et de la solution apportée</i>	20
IV.1. Une solution plus claire	20
IV.2. Gérer tous les périodiques	20
<i>V. Suite du travail</i>	21
V.1. Les "sans ISSN"	21
V.2. Le miroir à Nantes	21
<i>VI. Connaissances et expériences acquises au cours du stage</i>	22
VI.1. Le langage Perl	22
VI.2. MySQL	22
VI.3. CGI	22
VI.4. Les outils	23
VI.5. Reprise de programmes	23
CONCLUSION	24
BIBLIOGRAPHIE	25
ANNEXES	26

## **Introduction**

La Cellule MathDoc est une cellule documentaire nationale qui a été créée dans le but de faciliter l'accès à la documentation scientifique mathématique.

Elle ne possède aucun document mais met en place différents services permettant l'accès via le web à des ressources documentaires diverses.

Un des services proposés est le catalogue fusionné des périodiques de mathématiques, application développée à partir de 1995, devenue de plus en plus complexe au fil des années.

Mme Elizabeth Cherhal, ingénieur de MathDoc, souhaite pouvoir gérer ce système plus facilement. Elle m'a proposé un stage dont le but serait de reconstruire la base de données.

# **I. Contexte dans lequel le travail a été effectué**

## ***1.1. Présentation de la Cellule MathDoc***

### **Structure :**

La Cellule MathDoc (CMD) – Cellule de Coordination documentaire Nationale pour les Mathématiques – figure au contrat d'établissement 1999-2002 de l'UJF et a commencé à fonctionner, avec des moyens partiels, en décembre 1995.

Les missions confiées à la Cellule MathDoc portent :

- d'une part sur le dispositif de documentation recherche pour les mathématiques (coordination, soutien technique) avec pour objectif de faciliter l'accès à l'information scientifique et sa diffusion (catalogues de bibliothèques, prépublications, autres documents d'intérêt général),
- d'autre part sur la coopération franco-allemande dans la perspective d'extension européenne de la banque de données MATH-Zentralblatt (pilotage de la partie française de cette coopération, avec des aspects techniques, éditoriaux et structurels).

### **Moyens :**

Pour accomplir ces missions, la CMD dispose de personnels, de locaux et de moyens financiers.

#### **● Personnels :**

- deux professeurs d'université (Pierre Bérard, UJF, Directeur, et Laurent Guillopé, Université de Nantes, Directeur Adjoint),
- deux ingénieurs, ITARF UJF (Elizabeth Cherhal, mon maître de stage et Claude Goutorbe),
- une secrétaire, CNRS (Monique Marchand).

**● Locaux :**

Trois bureaux mis à disposition par l'UJF, dans le bâtiment du CICG. La Cellule MathDoc a par ailleurs bénéficié de certains moyens communs du CICG (salles de réunion, salle machine, imprimante).

**● Budget annuel :**

- 160 KF provenant de la DR (dont 60 KF de vacances)
- 100 KF de soutien de base CNRS.

## ***1.2. Le Catalogue Fusionné des périodiques de mathématiques***

Pour développer ses actions, la Cellule MathDoc a mis en place un serveur d'informations :

<http://www-mathdoc.ujf-grenoble.fr/>


A partir de ce serveur on peut accéder au Catalogue Fusionné. L'objectif initial de ce projet démarré en 1996 est de pouvoir mettre à disposition un outil de localisation des périodiques se trouvant dans les bibliothèques de recherche mathématique. Une autre composante, apparue avec force récemment, est de cataloguer les titres disposant d'un site web organisé par leur éditeur.

En juillet 1999, toutes les bibliothèques de mathématiques recherche de quelque importance participent à ce catalogue (à l'exception de la Bibliothèque de Jussieu). Le service enregistre en moyenne une dizaine de sessions d'interrogation par jour ouvrable.

Chaque bibliothèque fournit une copie du catalogue de ses périodiques (avec mise à jour régulière, de l'ordre de l'année en pratique) dans le format qui lui convient, à condition que celui-ci soit structuré et contienne un champ ISSN (pour les périodiques qui en sont munis). La Cellule MathDoc a développé des convertisseurs (des programmes en Perl) qui permettent l'insertion de ces données dans le catalogue fusionné.

L'affichage reprend des éléments de la notice ISSN (en accord avec le centre ISSN), des sites w3 concernant le titre, un ISSN électronique et l'état des collections des bibliothèques possédant ce titre dans leur fonds. Par exemple si on interroge le Catalogue

Fusionné pour le périodique « Bulletin des sciences mathématiques et astronomiques » on obtient :



### Catalogue périodiques :

**Bulletin des sciences mathématiques et astronomiques** [1155-8431] [Notice ISSN](#)

- [fra] 1870 -- 1884
- *Titre abrégé* : Bull. sci. math. astron.
- *Lieu de publication* : Paris
- *Éditeur commercial* : Gauthier-Villars
- *Absorbé partiellement par* : Bulletin astronomique (1884) [[0572-7405](#)]
- *Repris partiellement par* : Bulletin des sciences mathématiques (Paris, 1885) [[0007-4497](#)]

ISSN en relation

---


*Disponible dans les bibliothèques suivantes :*

- **Grenoble [IF]**
  - *État* : 1 (1870) – 11 (1876) ; 2eme serie : 1 (1877) – 8 (1884)
  - *Cote* : U.59
- **Marseille [CIRM]**
  - *État* : 1(1870) – 1884

Etat des collections des bibliothèques

---

Retour à la [page d'interrogation des périodiques](#)

 [Accueil@mathdoc.ujf-grenoble.fr](mailto:Accueil@mathdoc.ujf-grenoble.fr)

Début 97 a été installé un catalogue des titres pour lesquels la maison d'édition offre des informations (éditoriales, sommaires, texte intégral ..) sur la Toile. Ce catalogue est alimenté par la Cellule MathDoc au gré des annonces éditoriales et ajouts d'utilisateurs. Consultable par ordre alphabétique, il contient aujourd'hui 600 références et chacune de ces notices renvoie éventuellement aux informations du catalogue fusionné.

Le catalogue fusionné ne concerne que les titres pourvus d'un ISSN : l'extension aux autres titres est prévue depuis longtemps et n'a pu se concrétiser. Même si cette extension n'est pas prioritaire, elle fait partie d'une valorisation des fonds anciens tout en facilitant leur localisation, elle est donc à réaliser.

### ***1.3. Une gestion complexe***

La base de données et les procédures qui implémentent le Catalogue Fusionné sont entièrement faites en langage Perl. La consultation sur Internet du Catalogue Fusionné des périodiques fonctionne bien. La gestion du catalogue est en revanche difficile. L'accès aux données (stockées dans des fichiers textes Perl) n'est pas clair. Les relations entre les divers ensembles de données (les bibliothèques inscrites au Catalogue Fusionné, les périodiques stockés, les URL des éditeurs) ne sont pas faciles à visualiser et donc à modifier. Le plus souvent le plus simple pour consulter la base de données faite sous Perl est d'aller regarder sur le site w3 de la Cellule MathDoc. Mme Cherhal souhaiterait avoir une meilleure approche de la base en réalisant quelque chose de plus simple et plus clair.

## II. Sujet proposé

Voici le sujet de stage tel qu'il m'a été proposé :

« Reconception et réalisation d'un système de base de données concernant les périodiques »

Les étapes :

- Etude du système actuel (un ensemble assez compliqué de programmes en perl).
- Proposition d'amélioration de l'existant ou de refonte totale (en collaboration avec les ingénieurs de la Cellule MathDoc).
- Si refonte totale, réalisation de celle-ci, et écriture de l'interface CGI d'interrogation via le web.
- Si amélioration, réalisation et réécriture du cgi d'interrogation.

Pour ce stage, il est indispensable de connaître le langage perl. »

## **III. Travail réalisé**

### **III.1 Cahier des charges**

#### **III.1.1 Étude de l'existant**

J'ai dans un premier temps étudié les programmes en Perl qui géraient la base. Ne connaissant pas le Perl, j'ai consulté des ouvrages sur ce langage et visité des sites web (les références sont dans la bibliographie).

La base des journaux est construite à partir de fichiers texte qui contiennent des notices ISSN déchargées à partir du CD-ROM de l'ISSN et grâce à la routine nommée « `ecrire_db.pl` ».

Les catalogues envoyés par les bibliothèques sont des fichiers textes, tous sous des formats différents (la Cellule MathDoc n'a pas imposé de format) qui sont lus par des routines Perl. Ces routines remplissent des tableaux qui constituent la base des collections des bibliothèques, stockée dans un fichier nommé « `mult.per` » (voir les annexes).

La liste des bibliothèques se trouve aussi dans un fichier texte ainsi que la liste des URL correspondant à chaque périodique.

Lorsqu'une bibliothèque envoie un nouveau catalogue, il faut modifier une ligne du fichier « `l2p_cat.pl` » pour distinguer le nom de la bibliothèque qui change son catalogue, puis lancer la routine « `l2p.pl` » qui refait entièrement le fichier « `mult.per` » en modifiant ce qui concerne la bibliothèque en question. Une série de tests sont lancés lors de l'inclusion du nouveau catalogue :

- les périodiques sans ISSN sont-ils vivants ? (tout périodique encore publié doit posséder un numéro ISSN),
- le numéro ISSN est-il juste ? (un numéro ISSN répond à certaines normes),
- le numéro ISSN est-il unique pour chaque périodique ? (il arrive fréquemment que les bibliothèques envoient des listes où deux périodiques différents a priori ont le même ISSN),
- le périodique est-il dans la base des journaux ? (il faut télécharger de nouveaux périodiques du CD-ROM ISSN),
- l'ISSN est-il toujours valable ? (certains ISSN sont annulés).

### III.1.2 La nouvelle base de données

La base de données créée sous Perl était devenue trop gigantesque et compliquée. Pour rendre l'application plus claire, l'idée a été d'utiliser un logiciel de gestion de bases de données. MySQL étant librement disponible sur Internet et déjà installé sur le réseau MathDoc, la nouvelle base de données a été réalisée sous cette application.

Pour l'implémentation de la base sous MySQL, j'ai visité des sites qui apprenaient à démarrer sous MySQL et fait des petites manipulations sur des bases d'essai.

Après discussions avec les deux ingénieurs de la Cellule MathDoc, nous sommes arrivés à une base de données contenant six tables. Cette base se nomme 'Catalogue\_Fusionné'. Sous MySQL on met en colonne les champs de la table (Nom, Numéro, etc.) et en ligne les données entrées dans la table. L'organisation de la base est représentée en page suivante. Les six tables sont :

- Périodiques : 41 champs et 5483 périodiques
- Bibliothèques : 3 champs et 17 bibliothèques
- Etats\_Périodiques : 5 champs et 7110 abonnements
- URL\_Périodiques : 4 champs et 391 URL
- Services\_Sommaires : 3 champs (+ 1 rajouté plus tard) et 3 services de sommaires en ligne
- Périodiques\_Services\_Sommaires : 3 champs et 2251 pages web de sommaires en ligne.

Il y a deux sortes de tables dans la base. Certaines contiennent des entités chacune unique dans la base, les colonnes de la table décrivant leurs propriétés. Ce sont les tables suivantes : « Périodiques », « Bibliothèques », « Services Sommaires » qui contiennent les entités indiquées par leur nom. Les trois autres tables décrivent les liens existants entre les trois tables d'entités. C'est pourquoi elles contiennent des champs communs avec ces tables. Par exemple, la table « Etats\_Périodiques » contient tous les abonnements de toutes les bibliothèques. Elle a deux champs « Clé\_Périodique » et « Numéro\_ISSN » qui correspondent aux mêmes champs que dans la table « Périodiques », et un champ « Numéro\_Bibliothèque » qui correspond au même champ que dans la table « Bibliothèques ». Une ligne dans la table « Etats\_Périodiques » contient la clé et l'ISSN d'un périodique auquel la bibliothèque de numéro

« Numéro\_Bibliothèque » est abonnée, et l'état de la collection dans cette bibliothèque est décrit dans les champs « Etats » et « Lacunes ».

Les clés des tables (c'est-à-dire l'indice qui est unique pour chaque donnée et qui varie uniquement en fonction de cette table) sont uniques, non nulles et auto-incrémentées dans la table. C'est le cas de la « Clé\_Périodique » dans la table Périodiques, le « Numéro\_Bibliothèque » dans la table Bibliothèques et le « Numéro\_Service » dans la table Services\_Sommaires.

Dans toutes les tables, les noms des champs débutent par des majuscules et sont séparées par des « \_ ». Les clés communes à différentes tables ont la même orthographe dans toute la base.

La table « Périodiques » donne une clé unique pour chaque périodique. Le numéro ISSN a déjà cette propriété, mais l'idée est de pouvoir introduire dans la base les périodiques anciens ne possédant pas d'ISSN. La table « Périodiques » contient plus de 37 champs en tout, concernant les titres en relation avec le périodique (la suite, les fusions éventuelles,..) et des informations concernant le périodique (langue de publication, première année de publication, éditeur commercial,..).

La table « URL Périodiques » contient tous les sites qui concernent un périodique du Catalogue Fusionné. L'URL peut être de trois types : URL de l'« homepage » du journal (J), URL de la table des matières du site (T), ou une URL d'un autre type (A).

La table « Périodiques\_Services\_Sommaires » contient tous les périodiques qui ont les sommaires de leurs numéros en ligne grâce à l'un des trois services de sommaires qui gèrent des journaux à caractère scientifique (l'Inist, le sSs – le service de sommaire de MathDoc et Bordeaux – qui va disparaître puisque MathDoc reprend son rôle). Ceux-ci sont inscrits dans la table « Services Sommaires ».

Bibliothèques		
Numéro	Nom	URL
12	Lille	<a href="http://www.lille.fr/bib/math">www.lille.fr/bib/math</a>

Périodiques		
Clé	ISSN	Titre
324	1542-9857	New Scientist

Etats Périodiques		
Numéro Bibliothèque	Clé Périodique	Etat Collection
12	324	10 (1984) - 11 (1986)

URL Périodiques		
Clé Périodique	URL	Type URL
324	<a href="http://Mad-Scientists.com">Mad-Scientists.com</a>	J

Périodiques Services Sommaires	
Clé Périodique	N° Service
324	2

Services Sommaires	
N° Service	Nom service
2	Inist

## III.2 Réalisation

### III.2.1 Réalisation des programmes Perl qui rempliront la base

Dans la première version chacun des catalogues des 17 bibliothèques était lu dans une routine différente (exemple : la routine « l2p\_nancy.pl », la routine « l2p\_ens.pl », etc.) puis inscrit dans la base. Mais globalement les 17 fichiers « l2p\_nomville.pl » peuvent se découper en deux ensembles : une lecture pour des fichiers catalogues « tabulés » et une lecture pour des fichiers catalogue « non tabulés ». En effet, la liste des périodiques peut être rentrée de deux manières : chaque ligne contient un périodique et les informations concernant ce périodique (ISSN, titre, état) sont séparées par des tabulations ou un autre caractère distinctif.

*Exemple :*

*New Scientist      1542-9857 10 (1984) - 11 (1986)*

Ou bien les périodiques sont séparés par un saut de ligne et chaque ligne contient une information.

*Exemple :*

*Titre : New Scientist*

*ISSN : 1542-9857*

*Etat : 10 (1984) – 11 (1986)*

J'ai regroupé la lecture des 17 catalogues dans une procédure Perl nommée « lecture.pl » (voir en annexe) contenant deux parties : lecture tabulée ou lecture non tabulée.

Cette procédure produit un fichier nommé « nomville.sql » qui va contenir, pour chaque périodique, les trois champs à rentrer dans la table MySQL « Etats Périodiques » : le numéro ISSN, l'état des collections, les lacunes de la collection .

La routine « lecture.pl » fait aussi des tests sur la justesse de l'ISSN (ISSN faux inscrits dans le fichier « nomville.testb »), relève tous les titres sans ISSN (inscrits dans le fichier « nomville.ccn ») et en particulier ceux qui sont encore « vivants » (tout périodique vivant doit avoir un ISSN : les périodiques ne respectant pas cette règle sont inscrits dans le fichier « nomville.testa »).

Le chargement de la base se fait ensuite par la procédure « Remplir Table EP » (EP = Etats Périodiques). Celle-ci commence par effacer tout ce qui concerne la bibliothèque en question dans la table « Etats Périodiques », c'est-à-dire l'ancien catalogue. Puis elle charge le nouveau catalogue dans la table « Etats Périodiques » en s'appuyant sur le fichier « nomville.sql » provenant de « lecture.pl » et sur la table « Périodiques » dans laquelle elle va chercher la clé de chaque périodique.

Cette procédure fait aussi des tests. Elle regarde pour chaque périodique si le numéro ISSN n'a pas déjà été entré. Si oui, elle inscrit dans le fichier « nomville.testc » le titre du périodique déjà entré sous cet ISSN, l'ISSN en question et l'état et les lacunes associés aux deux doubles. Elle ne rentre pas l'ISSN en double, mais concatène l'état et les lacunes des ISSN identiques.

Le deuxième test effectué consiste à regarder si la notice du périodique existe dans la base (c'est-à-dire dans la table « Périodiques »). Si ce périodique manque, il est inscrit dans le fichier « nomville.testd » qui indique les périodiques à télécharger du CD-ROM de l'ISSN.

Pour remplir la base sous MySQL, j'ai écrit la procédure « lire\_bib.pl » qui remplit la table « Bibliothèques » à partir de la liste des 17 bibliothèques participantes et leurs URLs, et la procédure « ChargerTablePSS.pl » qui charge la table « Périodiques Services Sommaires » à partir de fichiers envoyés par les services de sommaires et la procédure « Remplir\_URL\_Perio.pl » qui charge la table « URL Périodiques » à partir d'une liste d'URL valables.

La liste des URL variant assez souvent (certains URL ne sont plus valables, d'autres sont signalés comme intéressants par les utilisateurs du site MathDoc), j'ai commencé à écrire une procédure « Table\_URL.pl » qui permettrait à une personne ne connaissant pas MySQL de modifier (ajouter, supprimer ou corriger des noms d'URL) les adresses électroniques de la table « URL\_Périodiques ». Cette procédure lancerait des requêtes MySQL à la suite des demandes de l'utilisateur sans que celui-ci ait à piloter MySQL. Mais cette procédure n'est pas terminée à l'heure où j'écris ce rapport.

### III.2.2 Problèmes rencontrés en refaisant l'application

En chargeant la table « Périodiques » à partir du CD-ROM de l'ISSN, nous avons, dans un premier temps, choisi que pour chaque champ qui désignerait un périodique nous entrerions son numéro ISSN, qui est un bon identificateur. Par exemple, les champs « Suite De », « Devient », « Fusionne\_Avec » qui désignaient des périodiques en relation avec la clé étaient représentés par leur numéro ISSN. Puis, comme nous avons mise en place la « clé » unique pour chaque périodique contenu dans la base MathDoc, nous avons pensé que le mieux était de désigner les périodiques par leur clé dans notre base, l'ISSN servant juste d'indicateur. J'ai lancé une procédure « ISSNVersCle.pl » qui transformait tous les ISSN en clés. Le problème qui s'est posé est le suivant : il y avait dans les champs « Suite De », « Devient », etc. 825 numéros ISSN non répertoriés dans la table « Périodiques », donc n'ayant pas de clé. La moitié de ces ISSN « non répertorié » se trouvait dans le champ « Suite De » : en effet, dans ce champ se trouve des périodiques anciens qui sont gérés sous leur nouveau nom, ils sont absents de la base MathDoc. Cela risque de poser un problème pour les routines d'accès aux données dans l'interrogation de la base (il faudra traiter différemment les clés et les ISSN).

L'autre problème qui se pose lors de la conception de cette base est lié à l'objectif d'intégrer les périodiques sans ISSN. Il faut pouvoir identifier ces périodiques par un autre moyen. Le titre n'est pas un identificateur très sûr, les bibliothèques rentrent parfois le titre en l'abrégeant ou se trompent et plusieurs périodiques peuvent se retrouver avec un même titre sans être forcément identiques. Le numéro CCN<sup>1</sup> a été envisagé comme moyen d'identifier les périodiques sans ISSN (c'est d'ailleurs un champ de la table « Périodiques », vide pour l'instant). Les notices de ces journaux seront versés dans la base plus tard, après conclusion d'un accord entre l'ABES (gestionnaire du CCN et la Cellule MathDoc).

---

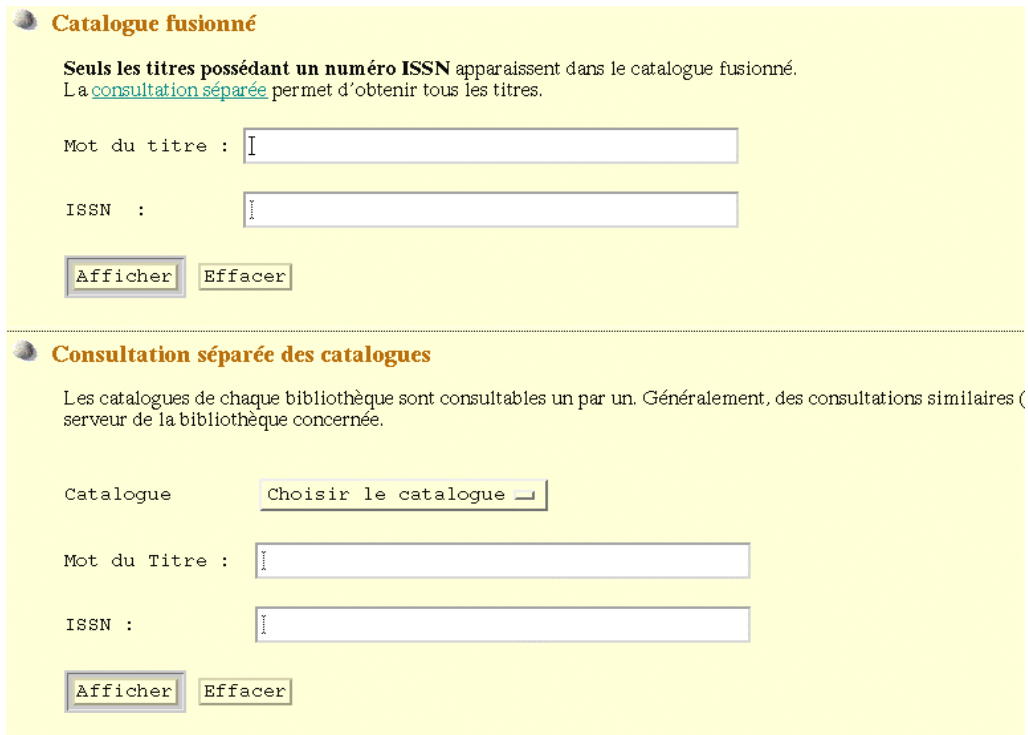
<sup>1</sup> CCN : catalogue collectif des périodiques. Voir <http://www.abes.fr/brev1.html>

### III.3 Implémentation du CGI

Il s'agit maintenant de revoir les routines qui interrogent la base et affichent les résultats de la requête sur le site w3 pour qu'elles ne travaillent plus avec l'ancienne base faite en Perl, mais avec la nouvelle sous MySQL. Le travail se situe plus dans des processus d'interrogation de la base à partir des mots-clés donnés par l'utilisateur que dans la conception de pages HTML. En effet, le site de la Cellule MathDoc est déjà fait et le passage à une nouvelle forme de base sera complètement transparent pour l'utilisateur (le site conservera la même mise en page).

#### III.3.1 Étude de l'existant

Comme pour la première phase, il s'agit de regarder comment l'interrogation de la base est faite actuellement. On peut consulter le catalogue des périodiques de deux manières : on demande le catalogue fusionné, qui recherche les périodiques dans toutes les bibliothèques inscrites, ou on sélectionne une bibliothèque particulière. Le formulaire d'interrogation se présente actuellement comme suit :



**Catalogue fusionné**

Seuls les titres possédant un numéro ISSN apparaissent dans le catalogue fusionné.  
La [consultation séparée](#) permet d'obtenir tous les titres.

Mot du titre :

ISSN :

---

**Consultation séparée des catalogues**

Les catalogues de chaque bibliothèque sont consultables un par un. Généralement, des consultations similaires (sur le serveur de la bibliothèque concernée).

Catalogue

Mot du Titre :

ISSN :

L'utilisateur entre, soit un mot du titre et/ou un numéro ISSN, soit un catalogue puis un mot ou un ISSN clé.

Dans la nouvelle application, le traitement pour le catalogue fusionné et pour la consultation séparée sera le même et les deux interrogations seront fusionnées. On pourra choisir dans les catalogues la mention « TOUTES » qui aura les mêmes fonctions que le catalogue fusionné. Ce regroupement devrait rendre plus clair la consultation des catalogues. Voici la nouvelle présentation :

**Catalogue des périodiques**

Seuls les titres possédant un numéro ISSN apparaissent dans le catalogue fusionné.  
La consultation séparée permet d'obtenir tous les titres.

Mot du titre :

ISSN :

Bibliothèque :

La page de réponse se trouve en page suivante.

Si on clique sur un titre, on accède à une notice plus détaillée du périodique, comme la notice se trouvant en page 5.

Pour l'instant l'interrogation des catalogues de périodiques fonctionne avec la routine « permul.pl ». Cette procédure recueille les requêtes de l'utilisateur, affiche la page des réponses à partir des données du fichier « mult.per » (dont il y a un extrait en annexe) ainsi que des pages d'erreurs pour guider l'utilisateur. J'ai, dans un premier temps, étudié ce fichier et déterminé ce que j'avais besoin de changer et ce qui pouvait rester intact. J'ai également lu et essayé des exemples de scripts CGI en Perl donnés dans les livres que m'avaient prêtés Mme Cherhal (voir la bibliographie).



## Catalogue périodiques : Nancy

Question : Mot du titre = theory

- **Ergodic theory and dynamical systems**
  - ISSN : 0143-3857 [Sommaires](#) (à Bordeaux)
  - État : 1 (1981) ->
  - Cote : Périodique: 099
- **IEEE transactions on information theory**
  - ISSN : 0018-9448
  - [Sommaires](#) (à Bordeaux)
  - [fas.sfu.ca](http://fas.sfu.ca)
  - État : 1963 (IT-9) ; 1974 (IT-20)-1 (1993)
  - Cote : Depot LORIA
- **Journal of approximation theory**
  - ISSN : 0021-9045
  - [Sommaires](#) (à Bordeaux)
  - Éditeur
  - [www.apnet.com](http://www.apnet.com)
  - [www.math.ohio-state.edu](http://www.math.ohio-state.edu)
  - État : 1 (1968) - 91 (1997)
  - Cote : Périodique: 121

### III.3.2 Réalisation

Pour réaliser l'interface CGI, j'ai tout d'abord créé la page d'interrogation en HTML (appelée « formulaire.html ») qui recueille les choix de l'utilisateur (les trois paramètres « mot du titre », « numéro ISSN » et « Catalogue »). J'ai repris la page d'interrogation actuelle en regroupant les deux sortes de catalogue, comme je l'ai expliqué plus haut. Dans le choix du catalogue, la valeur par défaut est « TOUTES » : on sélectionne alors les catalogues de toutes les bibliothèques.

J'ai ensuite travaillé à partir de fichiers d'essais et du fichier « permul.pl » pour arriver au fichier « my\_cgi.pl » qui a les mêmes fonctions que « permul.pl », mais interroge la nouvelle base sous MySQL au lieu du fichier « mul.per ».

Le programme principal de la procédure « my\_cgi.pl » appelle une dizaine de sous-routines assez courtes. Dans un premier temps, il cherche tous les titres contenant le mot entré par l'utilisateur, si la case « mot du titre » a été remplie (la fonction LIKE '%expression%' de MySQL trouve tous les titres contenant l'expression régulière 'expression' ). Sinon le programme trouve le périodique ayant pour ISSN le numéro indiqué par l'utilisateur ou fait une recherche à la fois sur les mots du titre et l'ISSN si les deux cases ont été remplies. Une fois les titres trouvés, on lance une requête pour connaître les bibliothèques qui possèdent le périodique ou bien pour savoir si une bibliothèque particulière a ce titre, dans le cas où l'utilisateur a précisé la bibliothèque de son choix. Il recherche ensuite les URL associées aux périodiques sélectionnés précédemment. Enfin, il recherche les services qui éditent les sommaires du périodique en ligne sur le web. C'est là où j'ai eu besoin d'une information qui n'était pas dans la base « Catalogue\_Fusionné ». Il s'agit de l'adresse qui donne accès directement à la page du site du service de sommaires relative au périodique. Cette adresse est nécessaire pour faire le lien hypertexte. La table des « Services\_Sommaires » contenait un champ URL qui donnait l'adresse de la page principale de leur site. J'ai rajouté une quatrième colonne appelée « URL\_Page\_Périodique » (ayant les mêmes propriétés que la troisième colonne « URL\_Service » : chaîne de 255 caractères) qui permet d'accéder directement à la page qui concerne le périodique.

*Exemple :* pour le service de sommaire « Inist » la page générale était : [www.inist.fr/public/fre/constl.html](http://www.inist.fr/public/fre/constl.html)

Mais pour accéder directement à la page qui concerne un périodique particulier, l'URL est :

[http://services.inist.fr/cgi-bin/public/f\\_search?modele=fre/f\\_resu91.htm&reference=](http://services.inist.fr/cgi-bin/public/f_search?modele=fre/f_resu91.htm&reference=)

suivi du numéro ISSN du périodique.

J'ai repris les procédures concernant les pages de traitement d'erreurs de l'utilisateur dans "permul.pl", elles renvoient des messages d'erreurs lorsqu'il n'y a aucune réponse à la requête, lorsque l'utilisateur n'a pas formulé de critère de recherche, etc.

## **IV. Intérêt du stage et de la solution apportée**

### ***IV.1. Une solution plus claire***

Les programmes Perl qui gèrent la base des périodiques actuellement sont assez longs, assez complexes et assez peu commentés. La gestion de cette base est difficile que ce soit pour visualiser la base ou pour apporter des modifications.

En choisissant un logiciel de traitement des bases de données pour traiter le problème, ces manipulations seront plus faciles. La gestion sera plus facile et plus rapide.

J'ai essayé de bien documenter les programmes que j'ai réalisés, afin que les personnes qui les lisent prochainement ne soit pas confrontés aux mêmes problèmes que moi.

### ***IV.2. Gérer tous les périodiques***

Actuellement, la base des périodiques de la Cellule MathDoc gère tous les journaux ayant un numéro ISSN. Un des objectifs est de pouvoir intégrer à la base des journaux sans ISSN (ce sont souvent des journaux anciens, mais ils sont cependant recherchés par les scientifiques qui consultent la base MathDoc). C'est pourquoi, en refaisant la base, nous avons mis en place la "clé périodique" qui permet d'identifier de manière unique un périodique de la base.

## **V. Suite du travail**

### ***V.1. Les "sans ISSN"***

Les périodiques "sans ISSN" doivent être intégrés à la base. Peut-être le seront-ils à la fin de mon stage (qui n'est pas terminé pour l'instant). Pour cela, il faut pouvoir en faire une liste complète et les identifier formellement. Comme je l'ai déjà expliqué plus tôt, le titre n'est pas une très bonne référence. La base est prête à intégrer ces périodiques, les procédures qui la remplissent en Perl devront se développer un peu.

### ***V.2. Le miroir à Nantes***

La Cellule MathDoc a un miroir à Nantes. Le système en Perl actuel marche aussi bien ici que là-bas. Pour terminer complètement l'implantation de la nouvelle base il faudrait installer MySQL ainsi que tous les fichiers nécessaires sur le serveur de Nantes.

## **VI. Connaissances et expériences acquises au cours du stage**

### ***VI.1. Le langage Perl***

Ce stage m'a permis d'apprendre le langage Perl. Les domaines que j'ai touchés en Perl sont :

- la manipulation des fichiers : lecture/écriture, notamment grâce aux expressions régulières qui sont très développées en Perl (dans la lignée d'Unix).
- les structure de données : les hachages, les tableaux et les pointeurs.
- la manipulation d'une base de données (ici MySQL) à partir d'un fichier Perl.
- le CGI d'interrogation (dialogue avec l'utilisateur).

### ***VI.2. MySQL***

J'ai aussi appris à manipuler MySQL qui est un bon logiciel de base de données. J'ai réfléchi à la conception de la base. Il est important de bien savoir quelles sont les meilleures structures à utiliser pour représenter l'information, quelles relations vont exister entre les tables, et quelles clés sont nécessaires dans la base.

### ***VI.3. CGI***

J'ai créé des pages web interactives. Cela m'a permis de faire connaissance avec le HTML et avec la bibliothèque CGI en Perl.

## **VI.4. Les outils**

J'ai travaillé sous Unix sur un Macintosh qui possède un émulateur X. Certains caractères comme le '|' (pipe) ou le '\ (antislash), très utiles en Perl et sous Unix, sont difficiles à obtenir sous Mac (combinaison de 2 ou 3 touches). Mais j'étais très contente de travailler sous Unix, je pense avoir appris des nouvelles commandes que je n'avais pas utilisées pendant cette année universitaire.

## **VI.5. Reprise de programmes**

Pour ce stage, j'ai dû reprendre un système déjà existant, comprendre ce qu'il faisait et comment il le faisait. Pour cela, j'ai dû lire des programmes faits par d'autres personnes de la Cellule MathDoc et bien souvent déchiffrer les programmes pratiquement ligne par ligne, car il n'y avait aucune documentation. Ce n'est pas toujours facile de comprendre les intentions d'un programmeur lorsqu'on a pas, soi-même, écrit le programme.

## CONCLUSION

Je suis très satisfaite d'avoir fait ce stage.

Tout d'abord, j'ai pu suivre mon projet du début jusqu'à la fin. En effet, j'ai touché à tout : d'abord la reconception de la base, puis les programmes qui l'alimenteront et enfin l'accès à cette base en tant qu'utilisateur du serveur d'information de la Cellule MathDoc. Le stage a été évolutif et très intéressant pour moi.

De plus, j'ai été bien encadrée, par des personnes très compétentes, ce qui m'a beaucoup motivée. La Cellule MathDoc est un endroit de travail très convivial.

## **BIBLIOGRAPHIE**

### Ouvrages sur le Perl

Randal L. Schwartz & Tom Christiansen  
"Learning Perl"  
[O'Reilly & Associates, Inc Unix Programming]

Larry Wall, Tom Christiansen  
"Programmation en Perl"  
[Editions O'Reilly International]

### Ouvrages sur le CGI/HTML

Isaac Cohen  
"CGI/Perl et Javascript, création de pages HTML interactives"  
[Eyrolles]

Philippe Chaléat, Daniel Charnay  
"HTML et la programmation de serveurs"  
[Eyrolles]

### Sites web consultés

MySQL :  
<http://www.mysql.org/>

Beginning MySQL Tutorial :  
<http://devshed.com/server-side/MySQL/Intro>

Initiation au Perl :  
<http://ww2.ec-lille.fr/~seguy/français/perl/index.html>

## **ANNEXES**

- Extrait du fichier "mult.per"
- Exemple de manipulation de la base sous MySQL
- Liste des tables de la base
- Fichiers Perl : "lecture.pl" et "RemplirTableEP.pl"