

# Rapport de stage M1

Stage réalisé en binôme par **Tarik ANSARI** (filiale informatique) et **Jean MARTY** (filiale électronique) du *17 avril au 17 août 2006* à l'*Institut d'Etudes Supérieures de la Mode (IESModa)*, au Mexique.



# Sommaire

<b>CONTACTS.....</b>	<b>3</b>
<b>PRESENTATION DE L'IESMODA .....</b>	<b>4</b>
<b>CONTEXTE ET OBJECTIFS .....</b>	<b>5</b>
BIBLIOTHÈQUE NUMÉRIQUE.....	5
INTRANET .....	6
<b>LA BIBLIOTHEQUE NUMERIQUE .....</b>	<b>7</b>
PHASE D'ÉTUDE.....	7
ARCHITECTURES ET TECHNOLOGIES CHOISIES .....	9
<i>Organigramme de l'architecture physique de la Bibliothèque Numérique .....</i>	<i>9</i>
<i>Stations de numérisation .....</i>	<i>10</i>
<i>Organigramme des étapes de la numérisation .....</i>	<i>14</i>
<i>Traitement des pages .....</i>	<i>14</i>
<i>Solution d'archivage.....</i>	<i>21</i>
<i>Mise en place du serveur, modifications du réseau.....</i>	<i>22</i>
<i>Site Internet : système de consultation.....</i>	<i>24</i>
L'ASPECT MANAGEMENT ET RH DU PROJET .....	30
<i>Phase de recrutement des étudiants .....</i>	<i>30</i>
<i>Planification du travail des étudiants .....</i>	<i>30</i>
<i>Méthode de management .....</i>	<i>31</i>
RÉSULTATS.....	32
<b>L'INTRANET.....</b>	<b>33</b>
PHASE D'ÉTUDE.....	33
ETAPES DU DÉVELOPPEMENT .....	33
1. <i>Refonte du design et de l'ergonomie .....</i>	<i>33</i>
2. <i>Restructuration de la base de donnée.....</i>	<i>34</i>
3. <i>Architecture du site et stratégies de sécurité .....</i>	<i>36</i>
4. <i>Reprogrammation des pages serveur .....</i>	<i>37</i>
5. <i>Mise à test parallèle.....</i>	<i>41</i>
6. <i>Ajout de fonctions, transition opérationnelle.....</i>	<i>41</i>
RÉSULTATS.....	42
<b>BILAN DE LA FIN DU STAGE .....</b>	<b>43</b>
Tarik Ansari.....	43
Jean Marty.....	44

## Contacts

### **Tarik ANSARI**

*EFREI Promo 2007*

Actuellement à : Beijing

Email : [tarik.ansari.fr@gmail.com](mailto:tarik.ansari.fr@gmail.com)

Téléphone : +86 136 8148 8454

### **Jean Marty**

*EFREI Promo 2007*

Actuellement à : Paris

Email : [jean.marty@gmail.com](mailto:jean.marty@gmail.com)

Téléphone : +33 6 69 17 25 73

### **Mme Emmanuele de Román**

*Directrice de l'IESModa et maître du stage*

Actuellement à : México

Email : [emmnr@iesmoda.edu.mx](mailto:emmnr@iesmoda.edu.mx)

Téléphone : +52 5514 2594

## Présentation de l'IESModa



L'institut d'études supérieures de la Mode (Instituto de Estudios Superiores de Moda), inauguré en 1998, est la seule école de mode mexicaine à proposer une formation obligatoire de français dans son cursus ; et à disposer d'une implantation au sein de la Casa de Francia (Maison de France) ainsi que de partenariats avec l'Université de la Mode de Lyon.

C'est par ailleurs l'un des rares établissements dans son genre au Mexique dont les programmes sont reconnus au niveau universitaire par le ministère de l'éducation mexicain. Financièrement indépendante, elle fait figure de pionnière face aux autres écoles de mode de la ville de Mexico, la deuxième plus grande ville au monde.



Comptant actuellement environ 230 étudiants réguliers, l'école est train d'inaugurer un nouveau campus et le premier centre culturel de la mode en Amérique Latine.



Désireuse de toujours garder une longueur d'avance dans tous les domaines vis-à-vis de ses concurrentes, l'école avait fait appel à un étudiant de l'EPITA l'an passé afin de réaliser un Intranet.

Un an plus tard, l'école veut aller plus loin et ouvrir une bibliothèque numérique et améliorer l'Intranet afin de prendre en compte un an d'expérience d'utilisation.

L'offre de stage de l'école avait été transmise à M. Huguet par le précédent stagiaire de l'EPITA ce qui nous a permis de la trouver dans les annonces affichés à l'intérieur de l'EFREI fin 2005.

## Contexte et Objectifs

Lors de notre arrivée, Mme de Román nous présenta César Olmes, consultant informatique qui se charge de la maintenance de son parc informatique. Son rôle fut de nous faciliter la tâche en nous donnant accès aux ressources informatiques de l'école dont nous avons besoin et de nous aider à contacter les fournisseurs mexicains.

### *Bibliothèque numérique*

L'école dispose de nombreux ouvrages du domaine de la mode dans la bibliothèque de la Casa de Francia. Les ouvrages de la mode étant spécialisés, ils sont généralement onéreux, par ailleurs nombre de ces ouvrages s'abiment, disparaissent, ou ne sont pas disponibles en quantité suffisante lorsque les élèves doivent faire des recherches. Afin de palier à ce problème, la constitution d'une bibliothèque numérique semble être la solution idéale pour la directrice de l'IESModa, et permettrait qui plus est, plus de flexibilité aux étudiants et à l'école.

Les objectifs qui nous ont été demandés sont le développement de la chaîne d'information de la Bibliothèque numérique (phase de numérisation, stockage, archivage, serveur et site de consultation), la conduite de la numérisation de 500 ouvrages, ainsi qu'une meilleure gestion possible des problèmes légaux vis-à-vis du droit d'auteur.

Pour faire la numérisation des ouvrages, l'école mit à notre disposition dès le début du stage deux étudiants effectuant leur service social (équivalent mexicain du stage français), qu'il fut à notre charge de superviser. Il n'y a pas de budget spécial prévu, mais l'école nous garantit de nous acheter le matériel nécessaire tant que ce sont de petites dépenses.

Les champs de compétences étant variés et le projet imposant, nous avons conduit ce projet en priorité et en binôme.

## ***Intranet***

L'école souhaite faire d'importantes améliorations sur l'Intranet à notre arrivée, surtout au niveau des fonctionnalités, mais aussi au niveau de l'ergonomie et du design. L'Intranet fonctionne sur un serveur dédié hébergé en Datacenter à l'extérieur de l'école et géré par César Olmes. L'école gère sur l'Intranet un certain nombre de services critiques comme la base de données des élèves, l'attribution, la consultation et l'impression des notes et une partie des informations délivrées aux élèves.

L'Intranet fonctionne en PHP (langage de programmation pour les pages serveur dynamiques) et fait appel à une base de données MySQL. Le serveur lui-même fonctionne sur le système FreeBSD.

Compte tenu qu'il s'agit d'un projet critique, sur lequel l'école requiert une certaine expertise afin de revoir éventuellement la conception de l'Intranet ; et que par ailleurs la Bibliothèque numérique requiert une grande partie de notre attention, Tarik Ansari prit en charge entièrement le projet de remise à jour de l'Intranet. Ceci, aussi compte tenu de son expérience en développement Web.

# La bibliothèque numérique

## ***Phase d'étude***

Nous avons tout d'abord commencé par faire une procédure de numérisation simple afin que les étudiants nous aidant puissent commencer à scanner les livres dès le début. Nous avons utilisé un ordinateur disponible dans la bibliothèque ainsi qu'un scanner à document A4 dont dispose l'école. Dans cette première étape, nous avons utilisé les logiciels fournis avec le scanner ou avec Windows et avons fixé la résolution de la numérisation à 300dpi (ce qui correspond à une résolution très détaillée), afin de pouvoir réaliser ultérieurement une reconnaissance de caractère sur les pages numérisées.

Pendant ce temps nous avons pu réfléchir aux différents problèmes et questions à résoudre, notamment, quel espace de stockage utilisent les pages numérisés en haute résolution, en quelle résolution les délivrer à l'utilisateur, quelles métadonnées (informations concernant le livre comme son titre ou son code ISBN) enregistrer pour chaque livre ; mais aussi comment tirer profit de ces métadonnées et du texte numérisé (« OCRisé ») pour l'utilisateur. En ce qui concerne la partie utilisateur, sur laquelle les étudiants et les professeurs doivent pouvoir consulter les livres directement en ligne, nous avons essayé de voir ce qui se fait déjà par le célèbre projet de numérisation d'ouvrages Google Books, afin de reprendre les fonctionnalités les plus appréciés.

Nous avons dut passer beaucoup de temps à optimiser la phase de numérisation, au fur et à mesure que nous avançons avec les étudiants. Un problème de manque de matériel s'est posé, ne disposant que d'un ordinateur pour cela, qui plus est pas assez performant pour les tâches que nous réalisons dessus. Nous avons donc essayés en vain pendant deux jours de récupérer des ordinateurs de l'école ne fonctionnant plus afin d'obtenir des ordinateurs supplémentaires. Nous nous sommes finalement chargés de commander deux stations DELL avec l'école afin de pouvoir avancer suffisamment rapidement.

Après plusieurs jours de « tâtonnements », nous avons pu établir les problématiques suivantes :

- Stations de numérisation : comment optimiser la rapidité de numérisation des pages et uniformiser l'interface de numérisation lorsque l'on utilise des scanners bon marchés de marques et spécifications hétérogènes ?

- Traitement des pages : comment convertir les pages aux formats requis et avec quel logiciels, comment régler les problèmes d'ombres à l'intersection des pages et recouper les pages le plus rapidement possible
- Numérisation du texte : quel logiciel OCR (Optical Character Recognition) utiliser, comment rendre cette étape la plus rapide possible
- Archivages des pages : que faire des pages le traitement terminé
- Serveur : de quel type de serveur avons-nous besoin et de quel système d'exploitation, sur quel réseau héberger le serveur, comment envoyer les pages des stations de numérisation au serveur
- Site Internet : quel architecture, quel type d'interface, comment résoudre les problèmes légaux, quels types de protections techniques utiliser

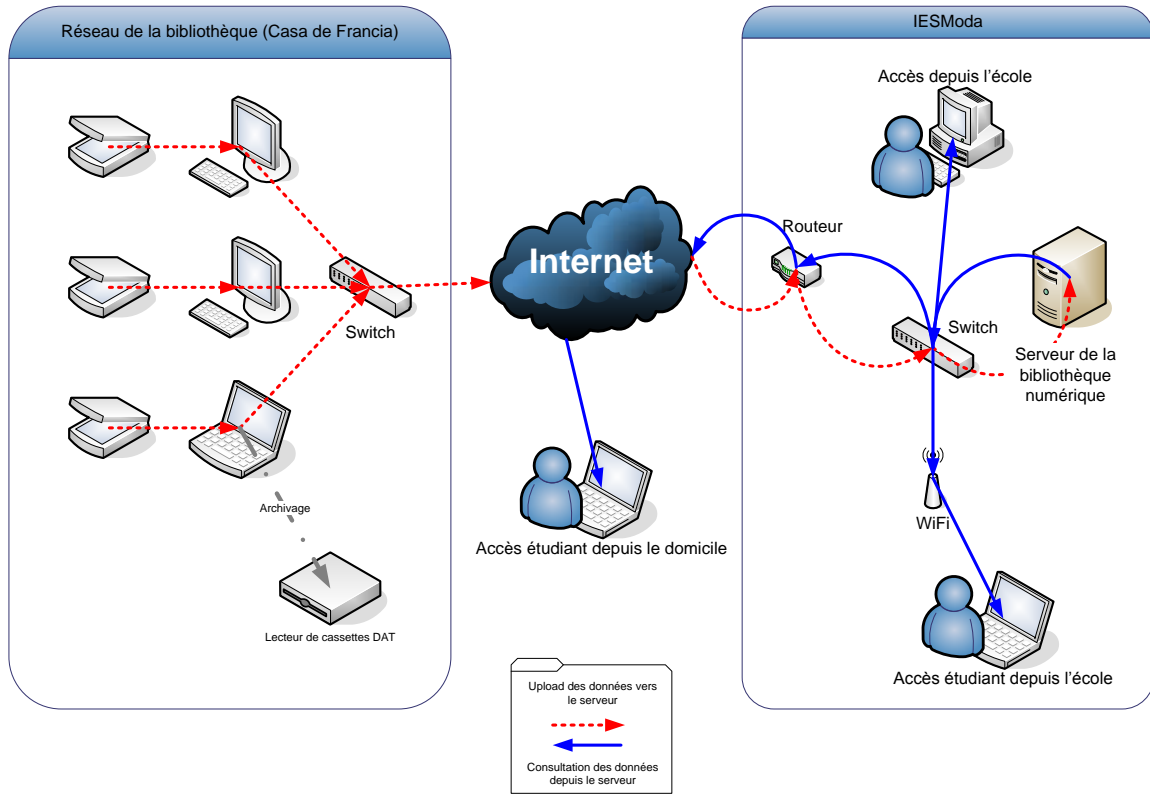
Avant la fin du premier mois, nous avons déployés une troisième station de numérisation utilisant un ordinateur portable de l'école. Egalement en manque de personnel pour numériser les livres à une cadence suffisante, nous avons recrutés en tout 10 étudiants n'ayant pas encore fait leur service social dans l'école. Nous les avons formés au processus de numérisation et avons fait un planning de façon à ce que les stations fonctionnent de façon optimale. Cela ne suffit malgré pas, nous nous sommes rapidement rendu compte que l'objectif de numériser 500 livres était difficile à atteindre au fur et à mesure que des nouveaux livres de plus 500 pages à scanner nous était fournis. Nous avons donc fait un rapport à la directrice lui expliquant qu'il nous fallait redéfinir cet objectif avec notre expérience du terrain, et en termes de pages et non de livres, nous avons alors proposé un nouvel objectif de 30,000 pages. Nous nous sommes basés sur l'estimation qui vient :

*Scanner une page prend environ 20 secondes sur les scanner à plat, soit 180 pages par heures, dans la pratique ce sont 100 pages par heure. En ajoutant le temps de traitement automatisé et manuel, l'OCRisation ainsi que l'upload vers le serveur, on arrive à 3 heures de travail pour la digitalisation de 100 pages.*

*Nous avons déjà scanné 7000 pages et il nous reste seulement 30 jours ouvrables de stage pendant lesquels nous pouvons scanner. Nous pouvons travailler 6 heures par jour sur chacune de nos 3 stations avec nos effectifs, ce qui permet de digitaliser entièrement 600 pages par jour. Cela permet d'arriver à un total de 25 000 pages à la fin du délai. En considérant que nous allons par ailleurs former des recrues supplémentaires et automatiser d'avantage le processus, un objectif de 30 000 pages nous semble atteignable sans difficultés.*

## Architectures et technologies choisies

### Organigramme de l'architecture physique de la Bibliothèque Numérique



## Stations de numérisation



Nos deux stations de numérisation dans la bibliothèque

### *Types de scanners*

Nous avons étudiés les différentes manières que nous avons pour scanner les livres et les traiter de la manière la plus automatisé possible (afin d'aller plus vite et de faire moins d'erreurs), nous avons notamment renseignés sur les stations manuelles de numérisation de livres disponibles sur le marché (beaucoup moins couteuses que les stations automatique, ou « robots »). Ces stations numérisent les livres à partir d'un support puis prennent des photographies hautes résolution allant de 200 à 300dpi, l'opérateur n'a qu'à tourner les pages et un logiciel spécialisé sépare la photographie en deux pages et coupe les zones noires. Ces stations présentent un intérêt considérable comparé à la numérisation de livre sur des scanners à plats notamment :

- Elles n'abîment pas les livres, alors que la numérisation à plat endommage considérablement les livres
- Elles permettent de numériser 10 fois plus de pages en un temps donné avec une seule station et un opérateur, contre 3 stations et 3 opérateurs dans notre plein régime
- Elles éliminent les problèmes de zones d'ombres, les photos étant faites à partir d'une certaine distance du livre
- Elles peuvent numériser les livres supérieurs au format A4 (nombreux dans notre bibliothèque)

Nous avons contacté Bookeye et i2S pour connaître le premier prix de leurs stations couleur manuelles scannant à 200dpi, et nous avons eu dans les deux cas une estimation approchant les 20,000 euros. L'investissement étant trop important pour l'école et surtout pour une utilisation qui ne dépassera guère 1000 livres, nous avons dû abandonner cette idée. Nous n'avons trouvé aucunes offres de location de ce type de scanner au Mexique ou au Etats-Unis.



**Kirtas BookScan 1200, une des machines utilisées par Google Books**

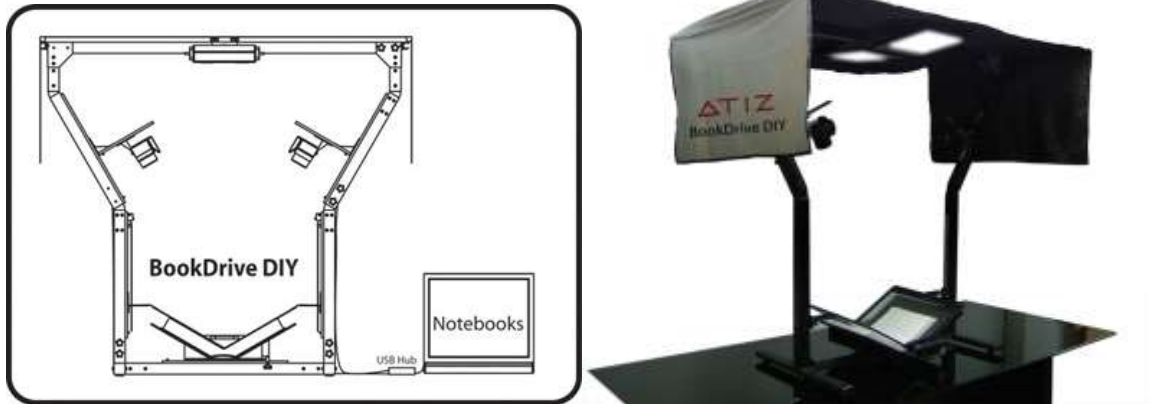


**Le CopyBook de la société française i2S**

Nous avons également étudié la possibilité de faire scanner les livres par Google dans le cadre de leur programme Google Library. Google disposent de robots coûtants plusieurs centaines de milliers de dollars capables de scanner et traiter de gros volumes en moins d'une minute. Nous avons dû également abandonner cette idée, la taille de la bibliothèque de l'IESModa étant trop petite pour susciter l'intérêt de Google, d'autant plus que la consultation des livres aurait posé trop de contraintes pour les élèves et de nombreux ouvrages n'auraient pu être consultables librement en ligne car ils sont trop récents.

À mi-stage, le projet ayant déjà bien avancé, nous sommes tombés sur toute nouvelle solution proposée par ATIZ (connue pour son scanner à livre automatique tenant sur un bureau) : le BookDrive DIY. Il s'agit d'une station de numérisation à faible coût sans capteurs optiques, simplement le support avec la connectique et les logiciels. L'utilisateur doit y installer deux appareils photo reflex numérique avec des lentilles à focale fixe, la station commande les appareils et les logiciels coupent les pages pour permettre de numériser 1000 pages à l'heure. Le coût de la solution est de 5000 dollars, 7300 dollars

en comptant l'achat des appareils photos reflex. Cette solution nous est apparue comme un bon retour sur investissement, l'école ayant alors aussi besoin d'acheter des appareils photos numériques professionnels. Malheureusement, l'école ayant alors déjà achetés deux ordinateurs supplémentaires et un lecteur DAT pour notre projet, la dépense est trop élevée. Nous nous sommes donc résignés à continuer à scanner sur des scanners à plat en optimisant au maximum les temps d'attente lors de la numérisation.



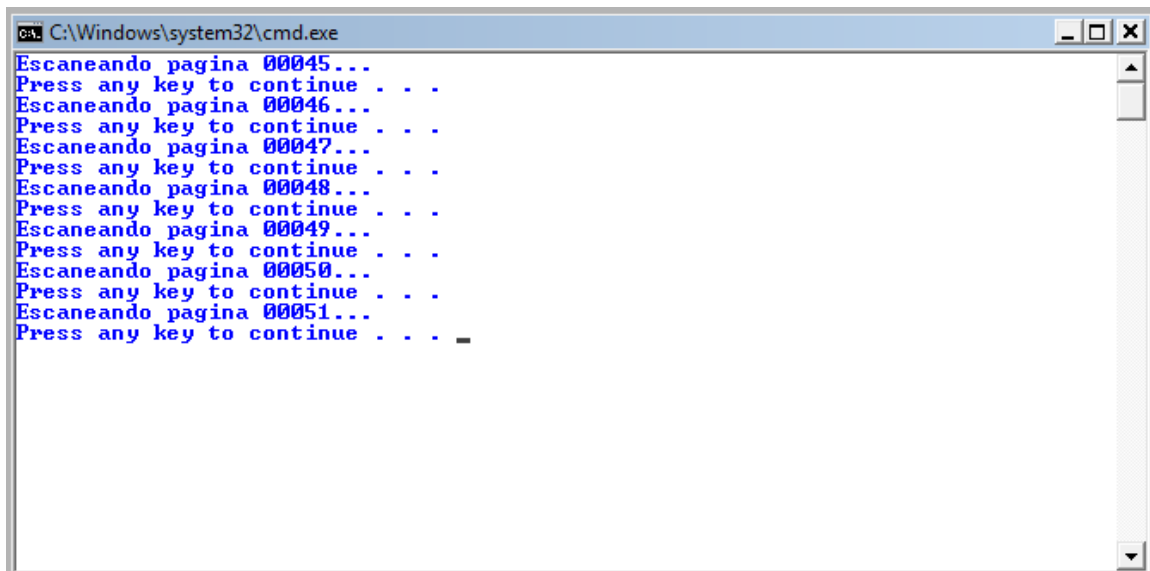
Par ailleurs, nous avons pensé à la possibilité de couper la reliure de certains ouvrages ou magazines disponibles en plusieurs exemplaires, et de les scanner dans le chargeur à document de notre scanner à plat HP, mais nous n'avons pas eu le temps lors du projet de numériser ce type d'ouvrage, la priorité ayant été donné aux ouvrage difficiles à trouver.

### ***Optimisation de la rapidité des scanners à plats***

Le problème se posant lors de l'utilisation de scanners pour le grand public est qu'ils font souvent appel à des drivers de mauvaise qualité ou tout simplement inadaptés pour scanner un grand nombre de pages répétitivement. Souvent l'interface du driver est trop détaillée et il faut redéfinir les paramètres à chaque page. En outre les logiciels de numérisation qui font appel au driver (ceux des constructeurs, de Windows, ou même les solutions de reconnaissance optique) utilisent une procédure qui ne convient pas, car après avoir scanné la page, ils bloquent le driver du scanner le temps d'enregistrer la page sur l'ordinateur. Nous avons donc testés un grands nombre de ces logiciels, et même des drivers TWAIN commerciaux, sans atteindre de résultats satisfaisants. Nous nous somme donc rendu à l'évidence qu'il nous fallait développer des drivers ou une interface pour les drivers par nous même pour obtenir des résultats corrects. Et c'est en cherchant des

librairies de contrôle de drivers de scanner de que nous sommes tombés sur une catégorie de programmes en ligne de commandes pouvant être scriptés pour contrôler les drivers d'un scanner. Nous avons alors comparé TwainSave de DTWAIN (89 dollars, 35 dollars pour le secteur de l'éducation), JSE Twain Commander (199 euros), Burrotech QuickScan (25 euros), Dosadi Acquire (99 dollars). Après avoir testé la version de démonstration des quatre programmes nous avons choisi TwainSave, celui-ci étant le seul proposant une compression sans perte en format de sortie, ce qui permet de gagner du temps d'occupation processeur sur les stations en évitant de faire la procédure par la suite.

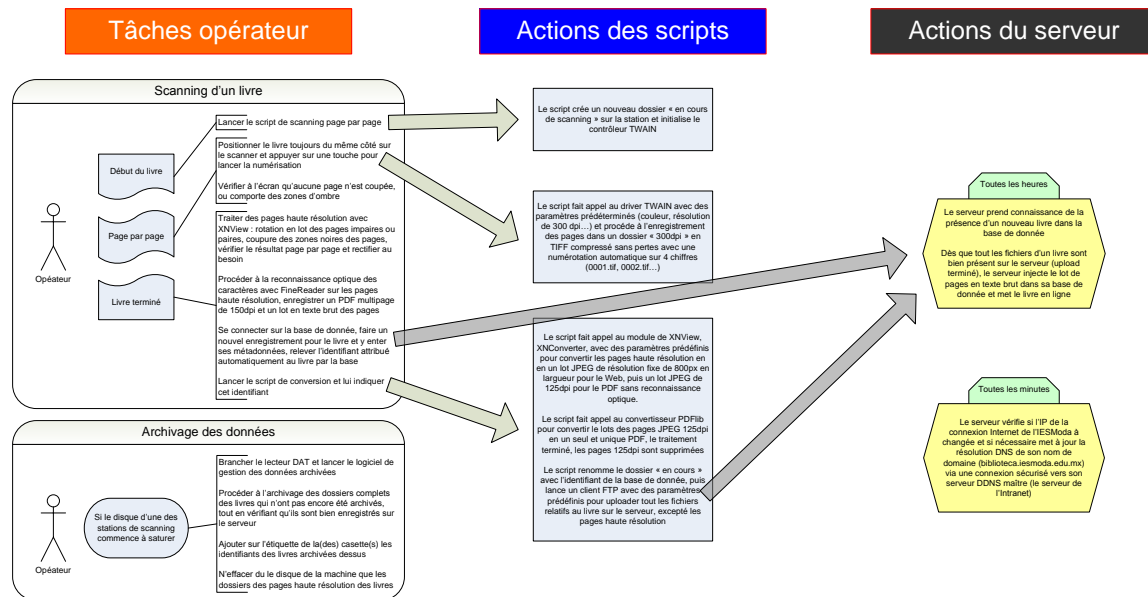
Nous avons alors intégré TwainSave dans un script Batch Windows écrit par nos soins qui se charge également de la création d'un dossier vide « en cours » pour le livre scanné, et de la numérotation des fichiers. Nous avons ensuite amélioré ce script pour qu'il puisse reprendre la numérotation d'un livre en cours après avoir été fermé. Le script terminé, l'opérateur dispose d'une interface on ne peut plus simple et scanne les pages trois fois plus vite qu'avec les solutions standards.



```
CA: C:\Windows\system32\cmd.exe
Escaneando pagina 00045...
Press any key to continue . . .
Escaneando pagina 00046...
Press any key to continue . . .
Escaneando pagina 00047...
Press any key to continue . . .
Escaneando pagina 00048...
Press any key to continue . . .
Escaneando pagina 00049...
Press any key to continue . . .
Escaneando pagina 00050...
Press any key to continue . . .
Escaneando pagina 00051...
Press any key to continue . . . -
```

**Interface simple mais script efficace**

## Organigramme des étapes de la numérisation



## Traitement des pages

### Étapes du traitement

Après avoir été scannées, les pages doivent être traitées pour éliminer les zones inutiles, améliorer la qualité (couleur et luminosité), puis doivent être converties dans les formats les plus appropriés pour l'archivage et la mise à disposition de l'utilisateur final. Nous avons décomposé ces étapes comme suit :

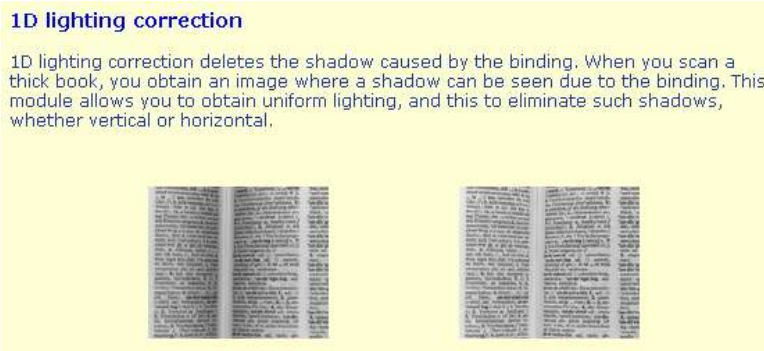
1. Prétraitement des pages : coupures des zones non voulues, élimination des zones d'ombre, calibration des couleurs et de la luminosité
2. Conversion des pages aux formats finaux
3. Archivages de la version haute résolution des pages

### Prétraitement des données du scanner

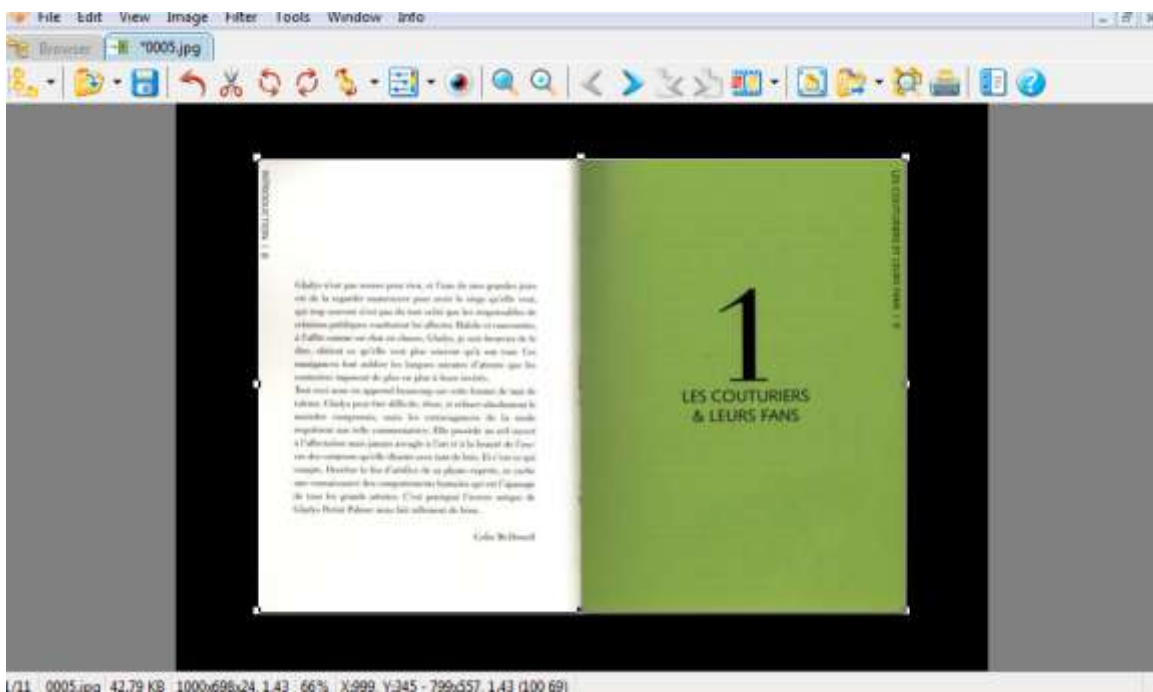
La calibration des couleurs et de la luminosité se fait en scannant une image modèle (comprenant des couleurs variées), puis en ajustant la valeur gamma (couleur) et luminosité du script de scanning jusqu'à obtenir à l'écran une image la plus fidèle à l'image du livre vue à la lumière naturelle. L'écran de l'ordinateur doit aussi avoir été correctement calibré auparavant pour que cette opération soit valide.

Nous coupons les pages à la fin du livre et avant la conversion des pages, pour cela deux méthodes sont possibles. La première consiste à couper les pages une à une à partir d'une interface comme XnView, cette opération prends environ 20 minutes pour 100 pages et a l'avantages de produire des résultats de bonne qualité. La seconde méthode consiste à observer le cadre sur lequel l'ensemble des pages du livre doivent être coupées et de faire une coupure automatique de toutes les pages selon ce cadre, opération aussi faisable sur XnView. La seconde méthode à l'inconvénient de « saboter » certaines pages étant donné que l'on scanne les livres sur des scanners à plats, l'alignement des pages n'étant alors pas toujours identique sur un même livre. Bien que cette méthode soit évidemment plus adaptée pour les livres scannés à partir de « scanners à livres », nous l'avons utilisé sur un certain nombre de livres bien scannés dans la mesure où cela ne posait pas de problème et pouvait nous faire gagner du temps.

Un des inconvénients majeurs lors de l'utilisation de scanners à plats et l'apparition de zones d'ombre sur la reliure des livres, zone qui s'étant parfois jusqu'aux paragraphes en marge du livre. Appuyer fortement sur le livre lorsque l'on le scanne ne résous évidemment pas convenablement le problème et abime encore plus les livres. Nous avons donc cherché des méthodes de traitement d'image pouvant faire disparaître ces zones d'ombre, dans un premier temps nous avons cherché l'existence d'un programme de traitement gratuit faisant cette opération, sans succès, nous avons alors essayé le logiciel de la solution commerciale de i2S disposant d'une fonction sensé réaliser le traitement de luminosité gradué devant résoudre le problème. Malheureusement nous avons été largement déçus par le résultat, le traitement rendant souvent la page encore plus mauvaise. Nous avons aussi cherché à faire ce traitement avec un script Photoshop sans trouver de solution fonctionnant convenablement et rapidement. En effet, l'étendue des zones d'ombres de livres scannés à plats varie selon les pages en raison de la variation de l'épaisseur du livre, nous en avons déduit que c'est un traitement trop difficile à effectuer avec les solutions actuelles pour un grand nombre de pages. Nous nous sommes résolus à scanner les pages en plaçant le livre au mieux sur le scanner et à couper ensuite les pages de la meilleure façon possible au niveau des marges.



### La présentation du logiciel Book Restorer d'i2S



### Le découpage d'une page à partir d'XnView

### *Choix de la résolution et des formats de compression*

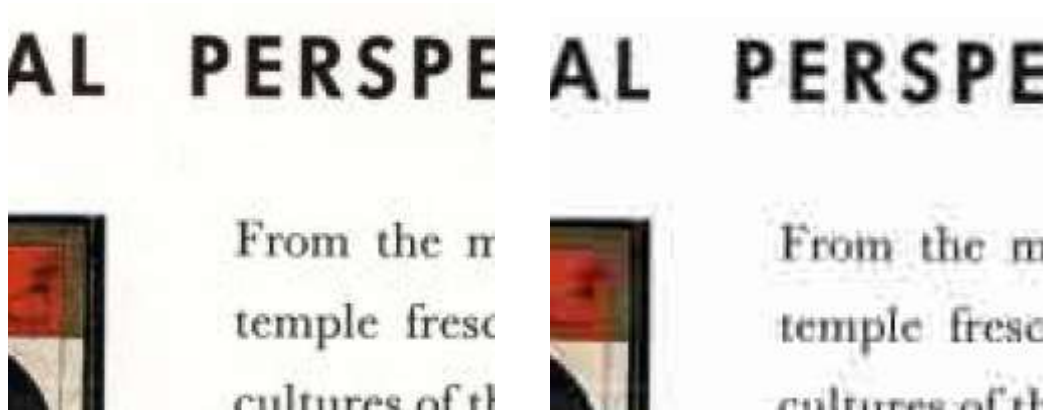
La résolution généralement recommandée pour scanner des pages et avoir ensuite la possibilité de réaliser une reconnaissance optique de caractère varie de 200 à 300 dpi (dots per inch). Nous avons choisi de scanner les pages en 300 dpi afin de pouvoir obtenir les meilleures résultats possibles et vu le faible coût de stockage qu'atteint le Gigaoctet. Une résolution de 250 dpi convient également très bien, et permet tout de même de reconnaître le texte des petites inscriptions (copyright, annotations), la majorité des scanners de à livres ne dépassant pas cette résolution. Au dessus de 300 dpi, la vitesse de la majorité des scanner à plat décroît considérablement.

Nous archivons ces pages « brutes » ou « haute résolution », au format TIFF qui permet une précision de couleurs de 24 bits et plusieurs algorithmes de compression sans pertes comme LWZ ou ZIP. La taille d'une page scannée en TIFF avec une résolution de 300 dpi varie entre 15 mégaoctets et 30 mégaoctets en fonction de la taille du livre. La compression sans perte permet ensuite de gagner entre 10 et 40% en fonction de la richesse en détails et couleurs de la page.

Viennent ensuite le choix des formats de sorties utilisés pour l'utilisateur final. Nous étudions dans cette partie des technologies de compression avec perte qui ont des caractéristiques de pertes de qualité variant selon l'algorithme utilisé. Les pages étant consultés sur le Web, nous avons cherché à minimiser au maximum cette taille avec une perte minimum de la qualité des pages afin de minimiser la bande passante utilisée par le serveur et d'obtenir des temps de chargement rapides pour l'utilisateur. Nous avons d'abord pensé au PNG pour sa compression n'entraînant pas de phénomènes de dégradation en flou sur le texte des pages, mais nous avons abandonné cette option, les pages étant généralement alors trop lourdes (> 200 Ko), nous avons préféré utiliser JPEG pour sa compression très forte et très bonne pour les images (les images occupent la plus grande partie des pages des livres de mode), les utilisateur souhaitant visualiser le texte très clairement, devant alors simplement ouvrir le livre dans sa version OCRisée. En ce qui concerne la résolution, nous avons choisie une résolution à largeur fixe (800 pixels) pour les pages directement visualisées sur le site Web, en raison de la méthode d'affichage des navigateurs qui ne redimensionne pas les images dans du contenu HTML. Nous avons choisit de redimensionner les pages des fichiers PDF en points par pouces, cette méthode est plus juste car elle s'adapte selon la taille réelle des pages. Nous avons pris une résolution de 125 dpi pour la version PDF non OCRisée, cela nous donne un fichier d'une taille d'environ 10 Mo pour un livre de 100 pages A4. Pour la version OCRisée des livres nous avons choisit une résolution de 150 dpi, le texte étant numérisé, la surface d'images à compresser est bien plus faible, ce qui autorise de les fournir dans une meilleure qualité. La taille obtenue du PDF OCRisée est d'environ 5 Mo pour un livre de 100 pages A4 contenant 50% de texte et 50% d'images.



**A gauche une image en 300dpi, à droite son équivalente en 100dpi**



**La compression JPEG rend mal les formes bien délimitées, comme le texte  
(image compressée à droite)**

### ***Logiciels utilisés pour la conversion***

Afin de réaliser ces opérations, nous avons scripté le module XnConverter de XnView et la librairie pdftk pour la génération des fichiers PDF.

### ***Considérations sur les performances de traitement***

Le traitement est une opération qui peut demander beaucoup de temps à l'ordinateur pour un livre si les étapes n'utilisent pas au mieux la mémoire et les opérations disque de l'ordinateur. Nous avons donc minimisé comme possible le nombre d'étapes avec les logiciels utilisés et les avons enchaînés dans un script. Cette opération dure environ 20 minutes pour 100 pages A4 sur nos machines (Pentium 4 2,4 Ghz). Lorsque l'ordinateur réalise la conversion, l'occupation CPU est de 100% et le système est difficilement utilisable pour scanner un nouveau livre, il peut donc être judicieux de trouver un moyen de limiter l'occupation CPU allouée au script appelant les programmes de conversion ou d'utiliser un processeur multi-core.

### ***Reconnaissance optique des caractères***

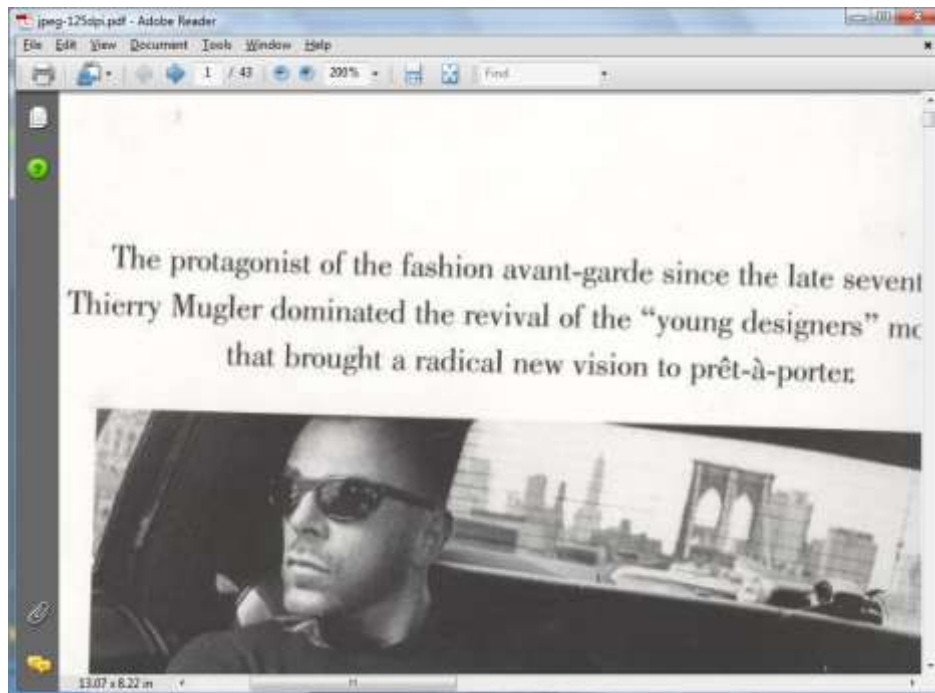
Un logiciel de reconnaissance optique des caractères se charge de décomposer une seule image (la page) en un texte associé à des illustrations.

Nous avons testés les performances des solutions majeures du marché : ScanSoft OmniPage, ABBYY FineReader, ReadIris, TextBridge. Dans notre cas d'utilisation performances signifient meilleure reconnaissance avec une action minimum de l'opérateur et rapidité du traitement. Bien que ce ne soit pas le logiciel doté de l'interface la plus évolué, FineReader nous a offert de bien meilleurs résultats que ses concurrents, généralement bien plus coûteux. C'est par ailleurs le logiciel fournit avec le BookDrive

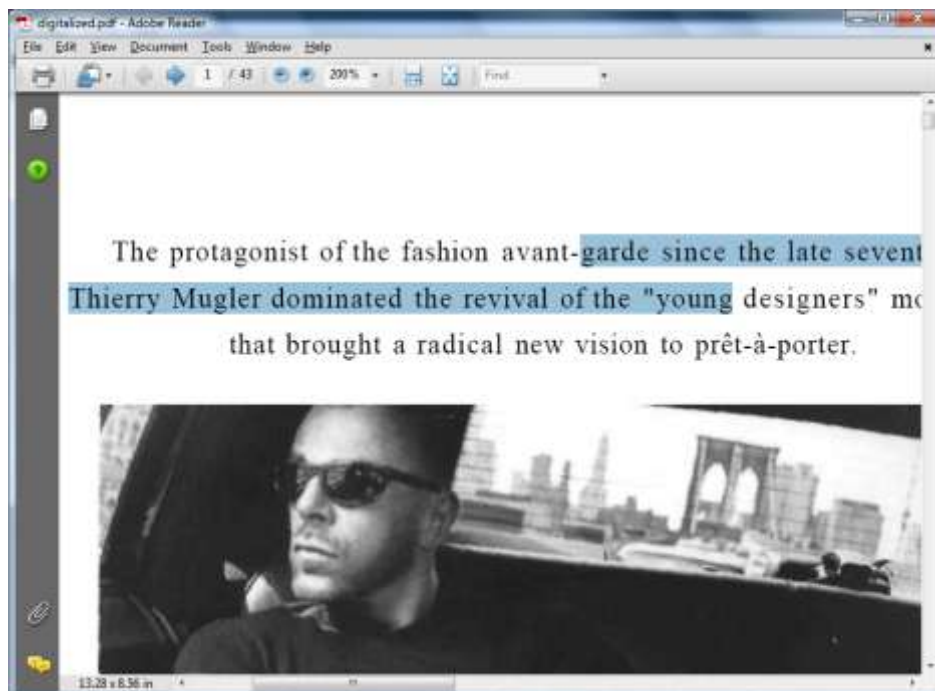
DIY de ATIZ. FineReader se charge directement de l'enregistrement du fichier PDF comprenant les illustrations sous forme d'images JPEG de 150 dpi, du code HTML des pages avec les illustration sous forme de fichiers images JPEG indépendants et du lot de pages en texte nu pour la base de données du serveur.

Une fois le texte d'un livre numérisé, le serveur est capable d'effectuer des recherches dans l'intégralité des pages du livre. Cela permet de demander au serveur tout les livres comprenant le mot « Robe » ou « Yves Saint Laurent », par exemple. Par ailleurs la consultation du livre avec son texte numérisé affiche un texte très clair quelque soit le niveau de zoom et offre plus d'interactivité à l'utilisateur, comme la possibilité de pouvoir sélectionner et copier un mot ou un paragraphe.

*Page suivante : comparaison d'un PDF avec et sans numérisation des caractères.*



PDF sans numérisation du texte (125dpi)



PDF avec texte OCRisé (images à 150dpi)

## Solution d'archivage



**Le StorageWorks DAT 72 de HP**

Les images brutes d'un livre occupent un espace de stockage très important (de 1,5 Go à 5 Go selon les livres), mais elles doivent être converties afin de pouvoir retraiter un livre, l'OCRiser plus tard, ou tout simplement faire des impressions, la grande partie des imprimantes actuelles imprimant en 1200 dpi. Afin de ne pas surcharger le serveur, et de contrôler les coûts nous avons opté de partir pour une solution de stockage périphérique. Nous avons d'abord étudié les offres de serveurs NAS (Network Access Storage), sachant que certains constructeurs disposent de prix très intéressants en raison de la baisse du prix des disques durs et de l'augmentation des capacités. Néanmoins, pour notre besoin, nous avons trouvé que l'utilisation de bandes DAT restait plus économique et souple, les pages brutes ayant très rarement besoin d'être récupérées une fois le traitement et l'OCRisation effectués.

Nous avons cherché un lecteur DAT pouvant écrire les cassettes DAT de 72 Go, externe et disposant d'une interface USB 2.0. DAT 72 parce que c'est la plus grande capacité disponible en technologie DAT (la technologie de bande la plus économique), et qu'une cassette DAT coûte presque le même prix quelque soit sa capacité ; Externe, afin de pouvoir déplacer facilement le lecteur d'un ordinateur à l'autre ce qui évite de surcharger inutilement le réseau ; l'interface USB 2.0, car elle est disponible sur tous les ordinateurs actuels, le débit de l'interface USB 2.0 étant supérieur au débit des lecteurs DAT 72.

Le produit le moins cher qui répondait à nos exigences était le lecteur d'HP (photo ci-dessus).

## Mise en place du serveur, modifications du réseau



**Le serveur de la Bibliothèque Numérique une fois installé**  
*En haut se trouve l'équipement réseau avec le modem ADSL et en bas les appareils de stabilisation électrique.*

### ***Besoins***

Nous avons tout d'abord établi que pour des raisons de flexibilité et de contrôle, la bibliothèque numérique devait avoir son propre serveur.

Ce serveur devait pouvoir stocker toutes les pages de la bibliothèque numérique, disposer d'une base de données des livres et des utilisateurs, d'un serveur Web pour délivrer les pages et d'un ensemble d'autres outils administratifs.

Un livre complet A4 de 250 pages dans ses différents formats occupant un espace d'environ 100 Mo sur le serveur, le serveur ne nécessite pas de capacités de stockage extraordinaires. Un disque de 40 ou 80 Go suffisant pour accueillir environ 400 à 800 livres moyens.

Un autre point très important concerne les besoins en puissance de traitement du serveur, la conversion des formats de fichiers des pages étant faite préalablement, nous faisons des économies de puissance de calcul importantes, le serveur ne générant pas les images en temps réel.

Pour finir, les besoins en bande passante sont quand à eux importants, le transfert d'images fait transiter en quantité les paquets sur le réseau, et il est important que les élèves puissent visualiser et télécharger les livres presque instantanément au moins depuis le réseau de l'école.

### ***Système et configuration***

Nous avons récupéré la première machine que nous utilisons dans la bibliothèque (Pentium 4, 512Mo de mémoire, 40 Go de disque dur). Nous l'avons formaté et avons installé dessus Ubuntu Server, qui est une distribution Linux. Au préalable, nous avons comparé différentes distributions Linux afin de voir laquelle pouvait nous permettre de déployer le plus rapidement notre serveur et le maintenir à jour avec le moins d'interventions possible.

Une fois l'installation terminée et SSH configuré (accès à distance au serveur), nous avons retiré l'écran et le clavier et avons placé le serveur sur le réseau. A partir ce moment, nous avons un serveur Linux disponible en permanence sur lequel nous pouvions installer des applications ou des services, changer les paramètres, télécharger des fichiers etc. Cela nous a donné la base pour développer le Site Internet et pouvoir ensuite uploader très rapidement les premiers livres.

### ***Hébergement et réseau***

Il n'y a pas au Mexique ou au Etats-Unis de Datacenter offrant de large bandes passantes mutualisées à de faibles coûts comme le propose OVH en France par exemple, il n'y a pas non plus de connections très haut débit symétriques bon marché comme ce que fait Free.

Nous devons pouvoir déployer le serveur rapidement, garder un contrôle de proximité dessus pour le développer et y avoir accès physiquement, l'héberger dans l'école s'est donc avéré une très bonne solution.

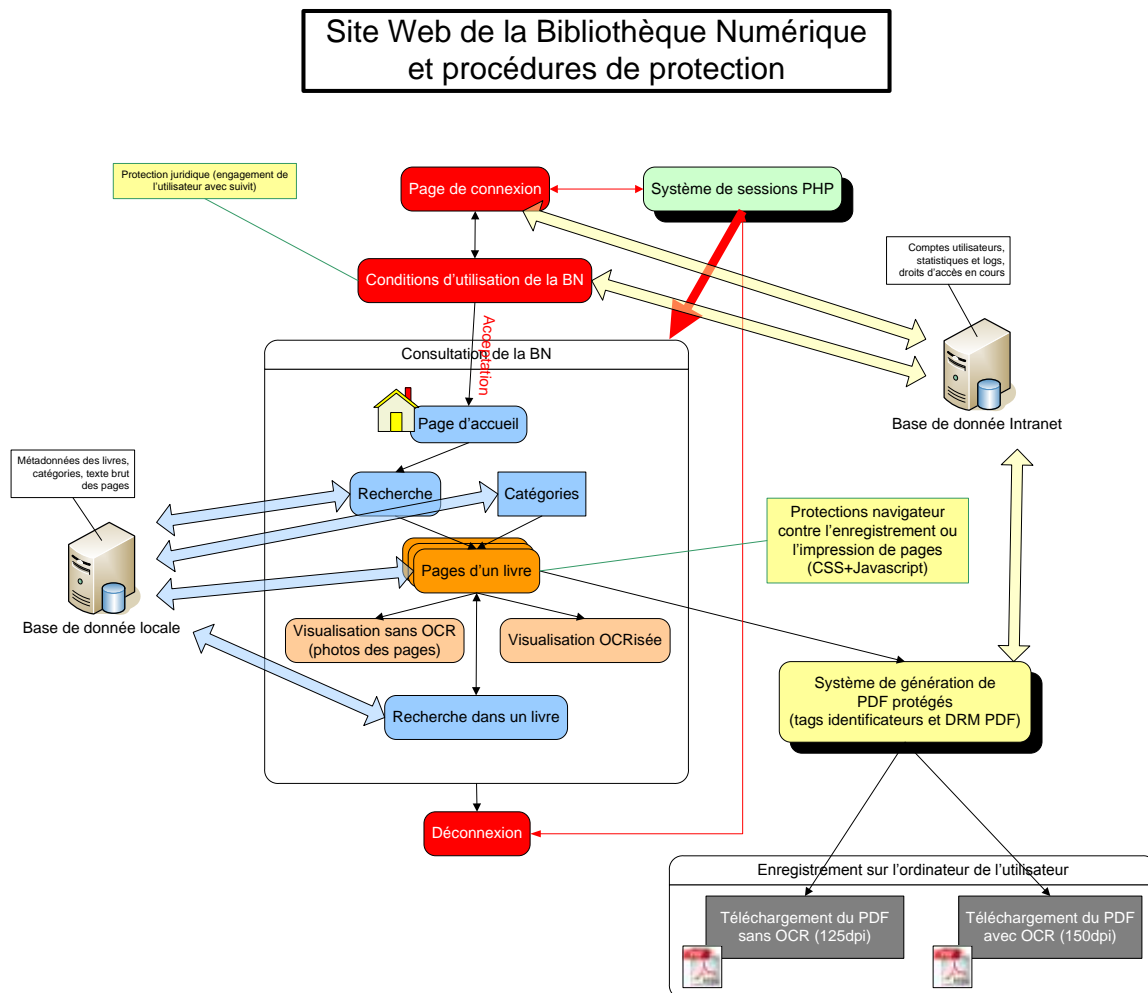
Le serveur étant hébergé dans l'école, la consultation est très rapide sur le réseau de l'école (100 Mbps), et il n'y a aucuns coûts d'hébergement, l'école disposant déjà d'une connexion ADSL de 1 Mbps.

Pour cela, nous avons du installer le serveur dans un espace aéré et protégé et nous avons déplacé la position du routeur et du Switch principal du réseau vers le serveur, afin qu'il n'y ait pas de nœuds d'étranglement au niveau du trafic réseau. Par ailleurs nous avons ondulé l'alimentation du serveur, de l'équipement réseau et de la connexion

Internet ; nous avons ensuite protégé le tout des surcharges et baisses de courant grâce à des appareils disponibles dans l'école.

## Site Internet : système de consultation

### Schéma de l'architecture



### Technologies

Afin que le service Web ne soit pas lourd en ressources pour le serveur, nous avons opté pour une architecture Web simple, et nous nous sommes appuyés sur un certain nombre de mécanismes satellitaires exécutés directement par le système d'exploitation (voir le diagramme « Actions du Serveur »).

Nous avons mis en place une base de données MySQL sur le serveur, celle-ci référence les livres avec le Titre, la Catégorie, le nombre de pages, la langue, l'éditeur, l'année d'édition, le numéro ISBN, mais également des champs booléens qui indiquent si le livre est entièrement uploadé sur le serveur et si il existe une version OCRisée. Les métadonnées sont remplies par les opérateurs et les champs booléens sont écrits par les scripts et le site se contente de les lire.

Nous avons choisit la technologie PHP, qui est adaptée à un site Web d'architecture simple et car nous avons déjà de bonnes compétences en PHP dans l'équipe.

Nous avons réalisé un design en XHTML/CSS, étape obligatoire quelque soit le site que l'on réalise. Nous avons réalisé un design simple en peu de temps avant de passer à la programmation PHP afin de simplement rendre dynamique un modèle de pages statique, c'est une méthode de développement qui nous a permis de collaborer dans le développement et d'avancer très rapidement.

### ***Interface d'accès et de consultation***

Nous nous sommes inspirés de certaines fonctions de l'interface de Google Books mais nous avons développé une ergonomie adaptée à un site qui regroupe des livres portant tous sur le même thème en réalisant des catégories et une visualisation des livres dans une définition plus importante. Par ailleurs, nous avons développé deux interfaces, une en français et une en espagnol, comme requis par la particularité de l'école.



Page d'accueil

Biblioteca Digital de IESModa - Mozilla Firefox  
 http://biblioteca.iesmoda.edu.mx/es/category/3

**Biblioteca Digital de IESModa**  
 Tu posición: [Página frontal](#) > [Historia del Arte](#)

Páginas digitalizadas: **21,639**

**Livros de Historia del Arte**

Desconectarse

Clasificar por: [Titulo](#) | [Autor](#) | [Fecha de publicacion](#) | [Editor](#) | [Fecha de digitalization](#) | [Numéros de paginas](#)

	<b>El Significado de los Colores</b> de Georgina Ortiz 0, Trillas 289 páginas - digitalizado por diana & grace		<b>Figure drawing for fashion</b> de Isao Yajima 1986, Graphic-Sha Publishing&Co., Ltd 134 páginas - digitalizado por Grace&Sandra
	<b>Histoire de la Beauté</b> de Umberto Eco 2004, Flammarion 441 páginas - digitalizado por Jean		<b>L'art du Tapa</b> de Emmanuel Anati 2005, Linsolite 195 páginas - digitalizado por Angie
	<b>Une Histoire de la Beauté</b> de Dominique Paquet		

Done

Biblioteca Digital de IESModa - Mozilla Firefox  
 http://biblioteca.iesmoda.edu.mx/es/book/10/5

**Biblioteca Digital de IESModa**  
 Tu posición: [Página frontal](#) > [Historia de la Moda](#) > [Comme des Garçons](#)

Páginas digitalizadas: **21,639**

**Comme des Garçons**  
de France Grand

◀ página 5 de 42 ▶

Desconectarse

Versión original (imagenes) | **Versión con texto digitalizado**

mode est rose à plus et ses connotations d'innocence ont à son tour, avec l'analyse intelligente de l'ethnologue (ce qui paraît très, très, très, remarquable à son propos est seulement ce qui est perçu) comme tel par nous culture à l'honneur inspirable de la jeunesse, mais pas à celle de la sexualité, comme incarnée au premier chef par les gens de mode. La sexualité n'est d'ailleurs pas tout de suite, surtout de l'effacement, dans la mode, dans la rue. Arrivent sur son histoire de mode glorieuse, l'Occident est tout à fait, tantôt perçue ou confusément devant ce monde imparfait.

Il se peut aussi que la mode en question des habitudes dans le domaine de goût, celui de la cuisine ou du langage par exemple, soit venue par la suite de la culture, partie de couleurs, de mots et de sons latentes ; ou des plus facilement "manger de saloir" - plus de ce qui pousse vers - ou "le fait tout simple" - au lieu de "le mode des modes au milieu de son histoire, que je suis une page."

**P**ourquoi la boutique française le salon français, la cuisine occidentale ? Pourquoi pas le salon de la rue, les gens de la rue ou l'échange au sein comme dans les lieux publics ? Pourquoi le shopping bag aller "sans faire" et pas l'approche en Arabie tout juste temporaire ? Et pourquoi l'innovation au début ou un instant pour lui des débuts ? Ce n'est et nous, si

la collection (France), appartenant à chaque époque Comme des Garçons par un regard facile qui nous fait à nous la marque elle-même, comme la pensée de votre regard et qui celle au présent. Chaque production, de l'habillement de parler sans être, comme un regard de celle dans son aspect de plastique, à commencer sans réalisation jusqu'à d'entraîner les parties qu'on a à un spectacle, sans être sans l'occasion d'entraîner ou se sera pas. Aller de l'avant sans être. Mais : comme une boutique qui se sera pas, ce qui est sans être sans usage du temps, notre regard, notre mode à nous, notre histoire.

**C**ette interrogation relative en forme d'éthnologie, de collections, de photographies s'adresse également à la pensée de mode : la pensée est sur le mode de ce que dit par Roland Barthes en 1967 : plusieurs photographies. Sans l'usage d'un objet, elle est "l'histoire de mode", puis "de la" (plans). La légende est-elle que le dit-elle. Les catalogues produits par Comme des Garçons dès le début des années quatre-vingt d'insistent cette redondance : "Le photo ne touche et je la serie de son habituel, "Ethnologie", "Bible", "Wagner", "Art", etc. ; ce n'est pas, faire les yeux, faire le détail remettre sur à la conscience affective." Chaque est un langage sans sans international. De cette pensée pour le photo et le graphique est la, véritable suggestion ancienne, sans obligation de précision, dans lequel Comme des

Portada  
 Derechos de Autores  
 Índice  
 Contraportada

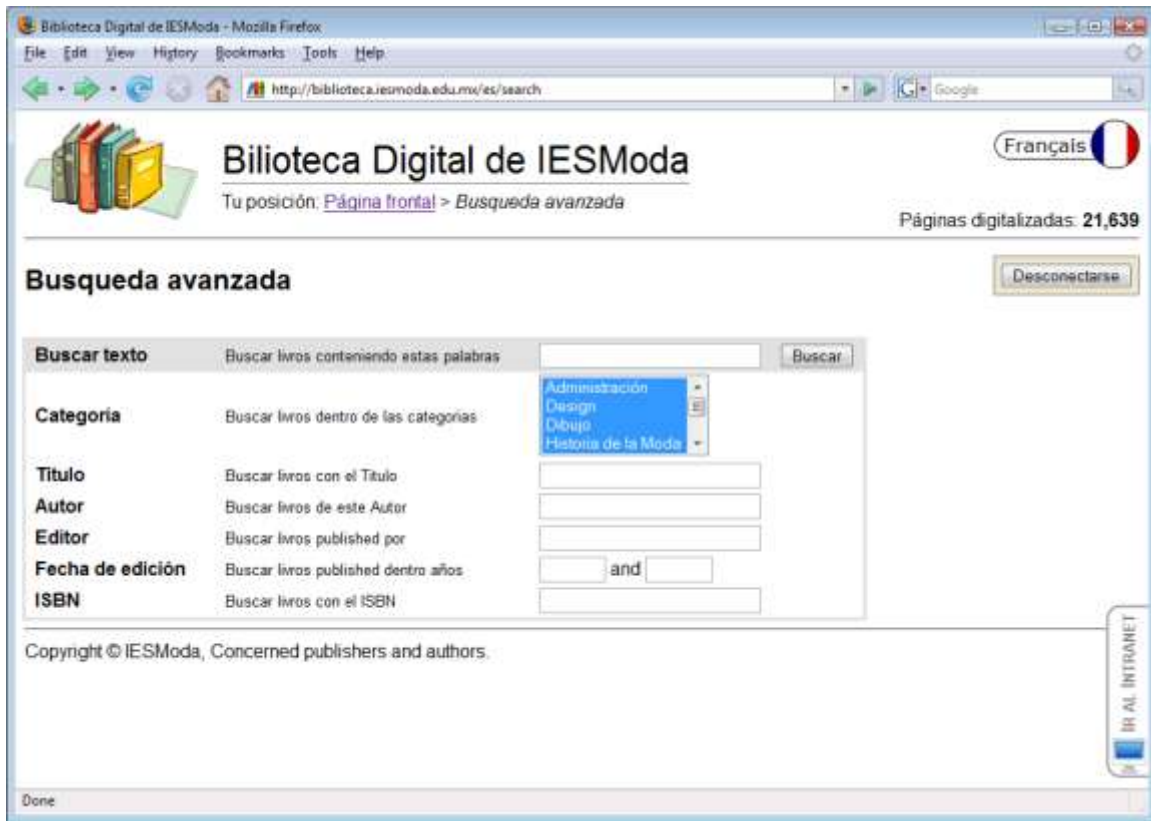
Buscar en este libro:

Editions Assouline  
 1998  
 ISBN : 2943230500

PDF original (imagenes) 18.3 Mo  
 PDF con texto digitalizado 18.4 Mo

Done

Nous avons ensuite réalisé un module de recherche de livre qui permet de faire des recherches de livre sur les métadonnées mais aussi sur le texte, grâce au texte des livres qui est inséré dans la base de données par le serveur après l'OCRisation.



### Recherche de livre

Ensuite, nous avons développé un module de session, afin qu'il faille s'identifier avec un nom d'utilisateur et un mot de passe pour accéder au site. Nous voulions que les élèves puissent accéder à la BN avec le même compte que le compte qui leur permet d'accéder à l'Intranet, afin de simplifier l'administration de la BN et l'accès aux élèves. Afin de pouvoir réaliser cela, nous avons dû faire dans le module de session une connexion SQL distance qui se connecte au serveur de l'Intranet pour donner accès au étudiant et enregistrer leur statistiques d'accès. Cela signifie que si le serveur de l'Intranet ne fonctionne pas, la BN n'est pas accessible, mais cela n'est en réalité pas un obstacle car le serveur de l'Intranet dispose d'un impératif de disponibilité bien supérieur au serveur de la BN.



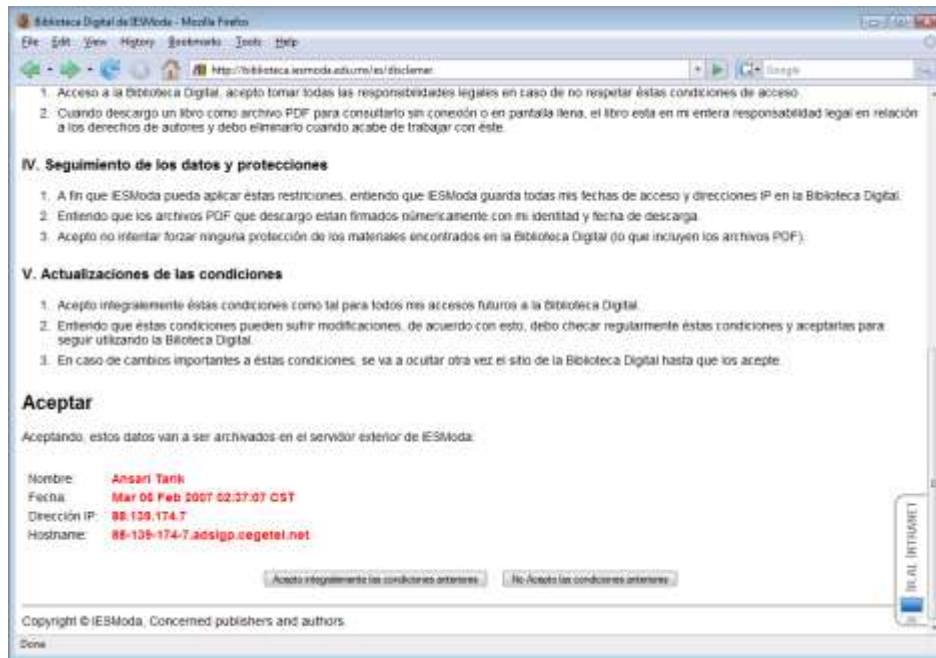
Page de connexion

### ***Mesures de protection des droits d'auteurs***

Il était impératif dans le cadre de ce projet que nous mettions en place toutes les mesures de protection possibles, légales et techniques, tout en permettant aux élèves d'avoir un accès complet à la bibliothèque.

D'un point de vu légal, nous avons dans un premier temps essayé de définir le cadre juridique de la BN face aux conventions internationales sur les droits d'auteurs. Afin de pouvoir donner accès aux livres au élèves sans surcoût tout comme ils le font dans la bibliothèque, nous avons établi qu'il s'agissait d'un cadre de consultation privé des livres constituant une simple extension de la bibliothèque. En partant de ce principe, il nous fallait mettre en place toutes les mesures possibles pour que l'accès aux livres reste exclusivement privé et que les personnes y accédant utilisent le site de manière strictement personnelle.

Nous avons donc écrit une chartre que tous les élèves doivent accepter avant de pouvoir utiliser la bibliothèque lors de leur première connexion. Une « signature numérique » de l'acceptation est conservée dans le serveur avec l'adresse IP d'accès la date et le nom de la personne ayant accepté la chartre.



#### L'acceptation de la charte est obligatoire pour pouvoir accéder au site

Par ailleurs, afin que les élèves ne puissent imprimer directement les pages depuis le site, nous avons mis une limitation CSS à l'aide d'une feuille de style pour l'impression. Cette mesure est facilement contournable mais adresse la méthode d'accès la moins sensible du site.

Les PDF sont en effet beaucoup plus sensibles, les élèves pouvant très bien les diffuser sur Internet après les avoir téléchargés sur leur ordinateur. Il fallait tout de même permettre aux élèves la consultation en PDF, qui est la plus ergonomique, et leur donne la possibilité de le télécharger un livre sur leur ordinateur pour le lire sans connexion Internet. Nous avons adressé d'une part le problème par des agréments dans la charte d'acceptation du site, et d'autre part de manière technique avec des DRM (mesures de contrôle des droits numériques) sur les PDF et un tampon numérique sur les PDF.

De manière pratique, lorsqu'un utilisateur veut télécharger un PDF, avant de lancer le téléchargement, le serveur verrouille la possibilité d'impression (fonctionnalité PDF), de copie du texte ou des images, et adjoint un « tag » dans le PDF qui contient le nom de l'utilisateur qui a téléchargé le PDF, la date, l'heure et l'adresse IP d'accès.

Si l'utilisateur diffuse un PDF (ce qu'il lui est interdit par la charte d'accès au site), il engage donc sa propre responsabilité.

Afin de pouvoir appliquer ce traitement au fichier PDF en temps réel sur le serveur, nous avons utilisé la librairie PDFlib avec PHP.

## L'aspect management et RH du projet

### Phase de recrutement des étudiants

Si nous avons bénéficié de la participation de deux étudiants dès le début du stage, nous avons ensuite présenté notre projet aux étudiants de l'IESModa à travers un PowerPoint. En effet, notre objectif de scanner le maximum de livre dans la bibliothèque numérique impliquait d'augmenter notre effectif. Nous avons argumenté que la bibliothèque numérique était pour eux, et que nous avons besoin de leur aide pour parvenir à la concevoir. Travailler avec nous leur permettrait aussi de valider leur service social, stage interne de l'école. Nous avons ainsi rassemblé une équipe de neuf étudiants, dont il a fallu assurer la gestion.

### Planification du travail des étudiants

	Lunes				Martes				Miercoles				Jueves				Viernes				Sabado			
	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
7h																								
8h																								
9h																								
10h	Diana	Graciela	Sandra	Cynthia	Diana	Graciela	Sandra	Cynthia	Diana	Graciela	Sandra	Cynthia	Diana	Graciela	Sandra	Cynthia	Diana	Graciela	Sandra	Cynthia				
11h																								
12h																								
13h																								
14h																								
15h	Paty	Angelica	Janet	Arianna	Paty	Angelica	Janet	Arianna	Paty	Angelica	Janet	Arianna	Paty	Angelica	Janet	Arianna	Paty	Angelica	Janet	Arianna				
16h																								
17h																								
18h																								
19h																								

Emploi du temps des étudiants

En concertation avec les étudiants, nous avons réalisé un emploi du temps hebdomadaire permettant un roulement continu sur les machines. De plus, nous avons tous l'obligation de pointer en arrivant et en partant sur un appareil qui relève les empreintes digitales. Les horaires de chaque stagiaire étaient ainsi stockés dans une base de données accessible à la directrice, avec un objectif, la ponctualité. Enfin, chaque étudiant détenait une fiche personnalisée avec le nombre d'heures travaillées et les tâches effectuées.

Pour augmenter la productivité des étudiants, nous leur avons enseigné toutes les étapes qui permettent de numériser un livre, et nous avons rédigé une procédure pour l'école « Adding a book inside the digital library ». Détaillée en anglais, langue technique traditionnelle, elle permet à un étudiant d'être rapidement autonome sur le travail à effectuer et lui fournit une vision d'ensemble du projet.

Les livres à scanner ont été listés par deux étudiants dans un fichier Excel. Chaque fois qu'un étudiant numérisait un livre et l'ajoutait dans la base de donnée de la bibliothèque, il rayait le nom du livre de la liste, et ajoutait son nom dans la base de donnée, de sorte que lorsque le livre s'affiche sur le site web, le nom de celui qui l'a scanné s'affiche aussi. Une simple requête dans la base de donnée de la bibliothèque permet ainsi de savoir qui a scanné quoi et quand.

### **Méthode de management**

La méthode de management que nous avons utilisé peut s'apparenter à du toyotisme. Afin de rendre le travail des étudiants attractifs et méritant, les étudiants réalisent tous l'ensemble des étapes de la digitalisation d'un livre. Lorsqu'ils ont terminé un livre, leur nom apparaît à coté du livre sur le site.

## **Résultats**

Le projet terminé nous avons digitalisé un peu plus de 20 000 pages et réalisé un serveur permettant aux étudiants d'y accéder de façon conviviale tout en respectant au mieux possible les droits des auteurs.

Sur la fin nous avons fait passer le délai de digitalisation pour 100 pages de trois heures à moins de deux heures, sans investir dans du matériel supplémentaire. Notre planning étant alors devenu très serré, et avec le projet de l'intranet demandant de plus en plus de temps, nous n'avons atteint que les 2/3 de nos objectifs en terme de nombre de pages disponibles lors de l'ouverture. Néanmoins, là n'était pas l'aspect le plus important de notre projet, l'essentiel de notre travail reposait sur la conception de la bibliothèque numérique et de la mise au point du processus de digitalisation.

Nous avons réussi à réaliser un système performant, le serveur étant à même de délivrer les livres moyennant une occupation faible de ses ressources et une utilisation modérée de la bande passante, même lorsque celui-ci délivre des pages à une vingtaine de clients simultanément.

Des améliorations sont encore envisageables comme la création d'un processus de digitalisation adapté pour la numérisation des revues de mode et une mise en avant qui leur est dédié sur le site. Cette petite idée semble un rien, mais permet donner une dimension actuelle au site, et donc pousser les élèves à consulter le site régulièrement. Le site peut alors jouer un rôle d'autant plus pédagogique en faisant là un pont entre l'actualité de la mode et ses classiques.

Avant tout, ce projet nous a démontré que les projets de bibliothèques numériques ne sont pas uniquement réservés aux grosses entreprises et aux gouvernements. Nous avons prouvé qu'il est possible de réaliser une bibliothèque numérique privée avec peu de ressources.

Aujourd'hui, si nous devions réaliser un projet similaire, nous insisterions néanmoins sur la nécessité d'investir dans une station de numérisation. Nous nous sommes en effet rendu compte que c'est une solution bien plus rentable que les scanners à plat dans la grande majorité des cas, et sur bien des points de vue.

*Nous avons travaillé ensemble sur le projet de la Bibliothèque Numérique, Tarik a coordonné les étapes techniques du projet et Jean a coordonné l'organisation des ressources humaines.*

# L'Intranet

*Projet réalisé par Tarik*

## **Phase d'étude**

La directrice m'a donné une liste de nouvelles fonctionnalités et rubriques qu'elle souhaitait voir apparaître sur l'intranet et m'a demandé d'obtenir une liste des bugs et problèmes rencontrés auprès de l'administration. Cela fait, j'ai commencé un état des lieux de l'intranet et de sa base de données afin de comprendre son architecture et de voir si celle-ci nécessitait également des améliorations.

Cette première étape m'a permis de trouver immédiatement plusieurs failles de sécurité sur l'intranet, ainsi que des imprudences majeures, comme le stockage en clair de mots de passe sur la base de données. Lors de l'analyse de l'architecture du site, j'ai trouvé de très nombreuses portions de codes se retrouvant redondantes dans les fichiers sources trop nombreux nommés une fois en anglais, et une fois en français ou en espagnol. Même constat au niveau de l'interface ; code html non valide, pas de séparation entre le contenu et le style, navigation faisant trop recourt à des JavaScript. Pour finir la base de données avouait elle aussi de nombreux défauts de conception : liaison des champs non fonctionnelle, mauvaise séparation des tables, identifiants non indexés...

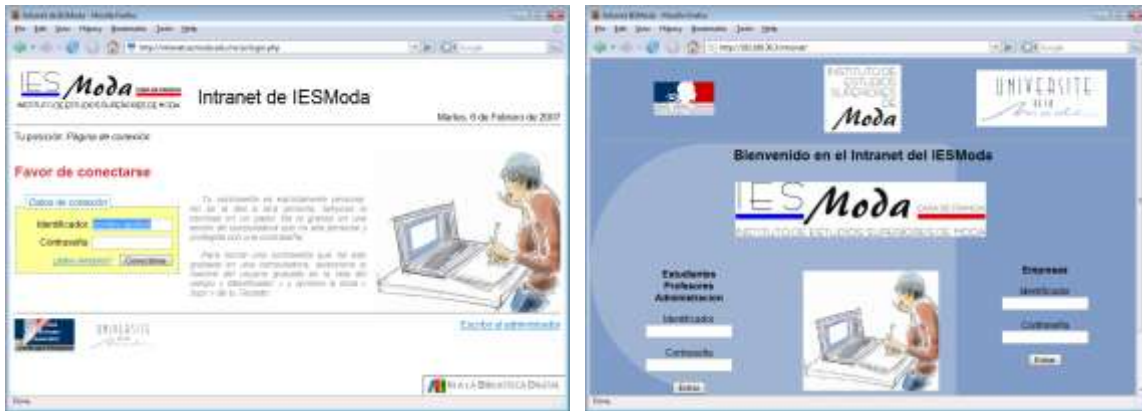
Il m'a donc semblé évident qu'une refonte de l'intranet ainsi que de sa base de donnée s'imposait avant de se lancer dans des améliorations fonctionnelles.

## **Etapes du développement**

### **1. Refonte du design et de l'ergonomie**

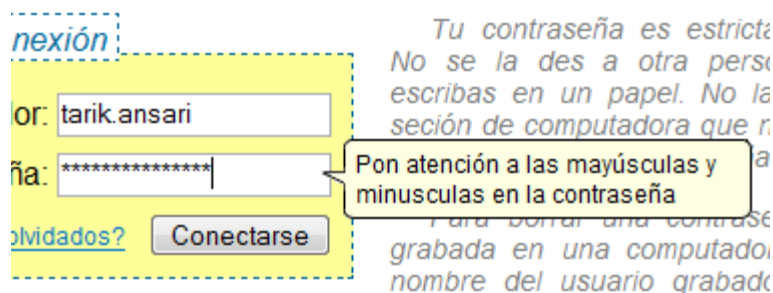
Alors que l'on travaillait principalement sur la bibliothèque numérique, j'ai commencé par faire une maquette pour le nouveau design de l'intranet. Le design comprend le codage HTML et CSS d'un site. C'est une étape essentielle car bien fait, il assure l'ergonomie d'un site, sa vitesse de chargement, ainsi que sa compatibilité auprès des navigateurs. Mais surtout, réaliser une maquette du design en première étape facilite grandement le travail de développement ultérieur en donnant un « cahier des charges vivant » permettant de voir en temps réel les résultats du travail.

J'ai concentré le design de l'intranet sur son ergonomie, partant du principe qu'une interface bien conçue doit permettre aux utilisateurs de comprendre son fonctionnement de la façon la plus naturelle et rapide possible.



Page de connexion du nouvel intranet / ancien intranet

La séparation du style (CSS) du HTML est une méthode presque obligatoire dans le développement Web d'aujourd'hui. De cette façon, toutes les informations de style du site (police de caractère, taille de police, couleurs, couleurs de font, bordures, positionnement, alignement...) sont contenues dans un seul et unique fichier, ce qui permet maintenir et de modifier le style d'un site de façon très aisée, sans modifier les autres fichiers sources. Cela permet aussi des économies de bande passante substantielles, le navigateur client n'ayant à télécharger ce fichier qu'une fois pour toutes les pages.



Aide contextuelle en CSS et JavaScript

## 2. Restructuration de la base de donnée

Il était essentiel de faire cette étape avant de reprogrammer le site ou de faire de nouvelles fonctions. J'ai donc repris la base de donnée en cours et ai vérifié la structure de ses tables unes à unes afin de supprimer les champs redondants ou non utilisés, tout en ajoutant les champs nécessaires aux nouvelles fonctions de l'intranet.

La plus grosse modification a été la refonte de la base d'utilisateur et du système d'identification à l'échelle de la base de données. Il y a trois types d'utilisateurs : élève, professeur et administrateur. Auparavant, chacun étaient sur une table distincte avec des champs redondants voir identiques mais non cohérents d'une table à l'autre. J'ai donc créé à leur place une seule table *users* sur laquelle se trouvent toutes les données communes aux utilisateurs dont le mot de passe chiffré avec l'algorithme SHA1. Un champ indique le type d'utilisateurs, et les données spécifiques à un type d'utilisateur se retrouvent sur les tables *u\_student* et *u\_prof*, les administrateurs n'ayant pas de données spécifiques.

Field	Type	Collation	Attributes	Null	Default	Extra
<input type="checkbox"/> id	int(10)		UNSIGNED	No		auto_increment
<input type="checkbox"/> login	varchar(20)	utf8_unicode_ci		No		
<input type="checkbox"/> sha1_passwd	varchar(40)	utf8_unicode_ci		No		
<input type="checkbox"/> user_type	set('student', 'prof', 'admin')	utf8_unicode_ci		No	student	
<input type="checkbox"/> access_active	enum('yes', 'no')	utf8_unicode_ci		No	yes	
<input type="checkbox"/> email	varchar(100)	utf8_unicode_ci		No		
<input type="checkbox"/> last_login	datetime			Yes	NULL	
<input type="checkbox"/> login_count	int(11)			No	0	
<input type="checkbox"/> forename	varchar(50)	utf8_unicode_ci		No		
<input type="checkbox"/> givennames	varchar(50)	utf8_unicode_ci		No		
<input type="checkbox"/> surname1	varchar(50)	utf8_unicode_ci		No		
<input type="checkbox"/> surname2	varchar(50)	utf8_unicode_ci		No		
<input type="checkbox"/> comments	tinytext	utf8_unicode_ci		No		
<input type="checkbox"/> sha1_parents_passwd	varchar(40)	utf8_unicode_ci		Yes	NULL	
<input type="checkbox"/> job	varchar(150)	utf8_unicode_ci		Yes	NULL	

Keyname	Type	Cardinality	Action	Field
PRIMARY	PRIMARY	774		id
login	UNIQUE	774		login
email	UNIQUE	774		email
access_level INDEX		9		user_type

**Structure de la nouvelle table *users***

Au fur et à mesure de la restructuration de la base de donnée, j'ai ajouté plus de précision à l'ensemble des définitions des champs afin d'augmenter le niveau de sécurité et d'intégrité de la base de donnée.

La dernière étape a été de transférer le contenu de l'ancienne base de données sur la nouvelle. J'ai programmé plusieurs script afin de réaliser cela, mais j'ai dû régler de nombreux conflits manuellement dut à des champs qui n'avaient plus de clef unique en passant sur la nouvelle base de donnée. Par exemple sur l'ancienne base de données, plusieurs utilisateurs pouvaient avoir le même identifiant ! L'intranet faisait la différence pour savoir si c'était un élève ou un professeur en comparant le mot de passe !



Tables de la nouvelle base

### 3. Architecture du site et stratégies de sécurité

La première étape a été de concevoir le site de manière à pouvoir créer des traductions de l'intranet en plusieurs langues de façon aisée. L'ancien intranet mélangeait parfois le français et l'espagnol sur ses pages. J'ai donc fait un répertoire de séparation qui sépare les pages pour la traduction. Les autres données (images, feuilles de style, etc.) se retrouvent dans des répertoires communs. Dans le répertoire propre à la langue se trouve également un fichier de configuration définissant entre-autres la « locale » que doit utiliser PHP pour le format de la date, le séparateur décimal etc.

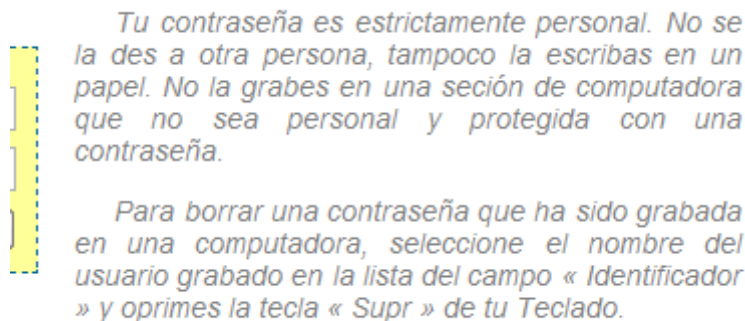
Name	Date modified	Type	Size
docs	6/26/2006 5:01 AM	File Folder	
es	1/22/2007 12:45 PM	File Folder	
fr	6/26/2006 5:01 AM	File Folder	
img	1/22/2007 12:46 PM	File Folder	
photos	1/22/2007 12:46 PM	File Folder	
style	1/22/2007 12:46 PM	File Folder	
.htaccess	8/8/2006 8:16 PM	HTACCESS File	1 KB
favicon.ico	8/16/2006 7:35 PM	Icon	2 KB
index.php	8/8/2006 9:39 PM	PHP File	1 KB
logout.php	8/8/2006 9:39 PM	PHP File	1 KB

Racine http de l'intranet

Le système de sessions a été programmé en utilisant les classes prévues à cet effet dans PHP, les cookies doivent obligatoirement être activés pour renforcer la sécurité (cela évite à un hacker de s'approprier une session grâce à la connaissance de l'URL visitée).

La session expire obligatoirement au-delà de 5 minutes de non utilisation. L'accès à toutes les pages de l'intranet est vérifié par une fonction d'en-tête et la variable *\$u\_type* permet de savoir à tout moment dans les pages le niveau d'accès de l'utilisateur authentifié : *student*, *prof*, ou *admin*.

La sécurité technique d'une application étant insuffisante en elle-même, l'accent a aussi été mis sur la communication à l'utilisateur : ne pas enregistrer son mot de passe dans le navigateur, ou divulguer son mot de passe sont quelques mesures simples mais pourtant vitales sur un site sensible comme un intranet qui inclus le système d'enregistrement des évaluations des élève.



#### Message aux utilisateurs sur la sécurité

Pour finir, une page permet aux utilisateurs ayant oublié leur nom d'utilisateur ou leur mot de passe de récupérer un nouveau mot de passe par email. Cela permet de limiter le nombre d'interventions nécessaires à l'administrateur et évite aux utilisateurs de noter leur mot de passe sur des bouts de papier.

Este formulario te generará una nueva contraseña por email. Tu contraseña nunca no se podrá recuperar.

E-mail:

[Regresar a la página de conexión](#)

#### 4. Reprogrammation des pages serveur

Les premières étapes terminées, la reprogrammation complète de l'intranet m'a pris deux semaines de travail, l'ergonomie a été revue pour la totalité des fonctions, en mettant l'accent sur la sécurité et la convivialité.



Page d'accueil

Les plus gros modules étaient le système de profil des utilisateurs (affichage et recherche d'un utilisateur), modification, ajout d'un utilisateur. Ces pages jouent sur un grand nombre de champs au niveau de la base de données sur lesquels le contenu doit systématiquement être vérifié. Un JavaScript permet de générer automatiquement les noms d'utilisateurs.



Listing des étudiants



Ajout d'un professeur

Le second gros module est le système de notation des élèves, après avoir créé les semestres d'un groupe et lui avoir affecté des cours, les professeurs remplissent les notes et les absences de leurs étudiants directement sur l'intranet. Les calculs des moyennes et des totaux sont entièrement automatique. Pour finir, une vue spéciale des listes d'élèves et des bulletins de notes ont été créés avec leur propre feuille de style pour l'impression.

**Table 1: Course Grades**

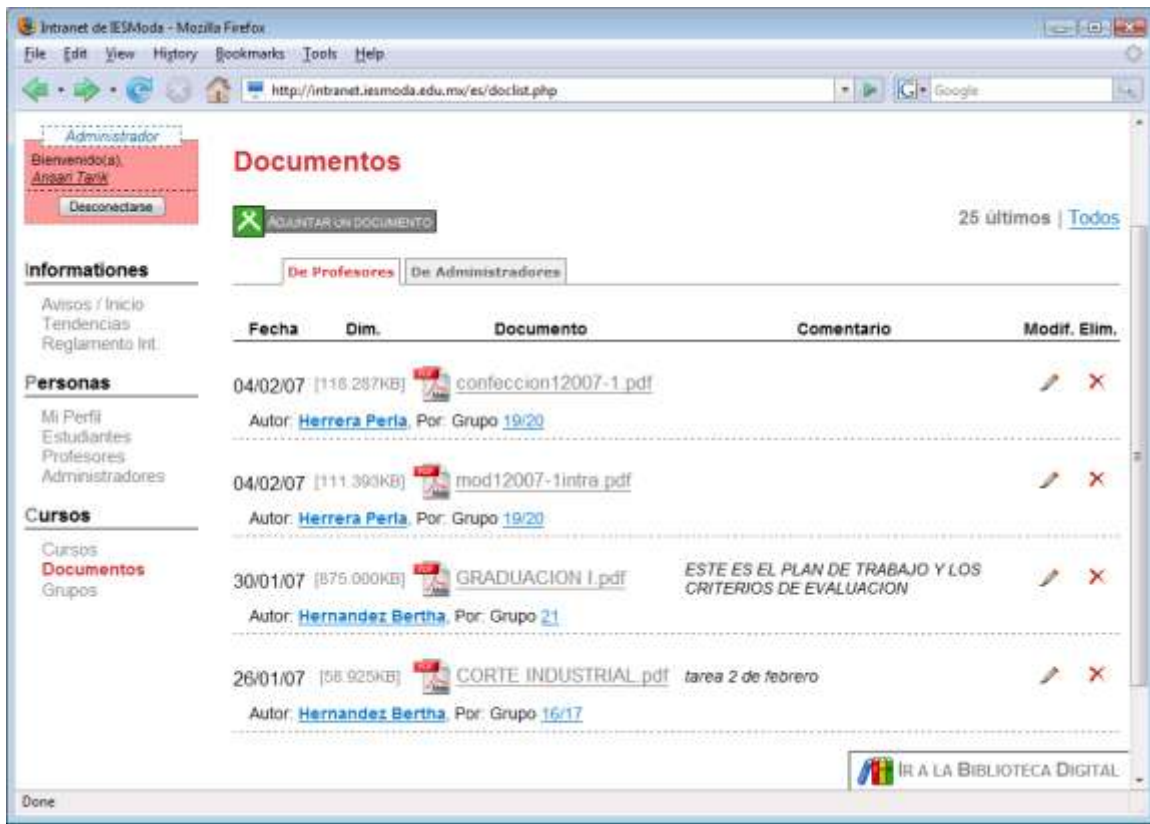
Curso	1 P	2 P	3 P	Final	Prom	1 P	2 P	3 P	Final	Total
Modelado Sobre Maniqui III (Patronaje) Henera Perla	8.2	8.5	9	8.8	8.6	5/0	4/1	4/1	3/0	10/2
Sistemas de Producci Hernandez Bertha	9	9	9	9	9	5/0	0/0	0/0	0/0	5/0
Corte Industrial I Hernandez Bertha	9	9	9	9	9	5/0	0/0	0/0	0/0	5/0
Graduaci Hernandez Bertha	9.4	9.6	9.8	9	9.5	5/0	5/0	4/0	3/1	17/1
Taller de Tendencias De Roman Emmanuele	8.6	8.6	8.6	8.6	8.6	5/2	4/0	4/0	4/0	10/2
Expresion PI Najera German	8	9	10	10	9.3	10/1	5/1	5/0	5/0	25/2
Taller de Accesorios Siva Jose	9	9	9	9	9	5/0	0/0	0/0	0/0	5/0
<b>Promedios / Totales:</b>	<b>8.6</b>	<b>8.8</b>	<b>9.2</b>	<b>9</b>	<b>8.9</b>	<b>52/5</b>	<b>28/3</b>	<b>25/2</b>	<b>22/1</b>	<b>127/11</b>

**Table 2: Semester 7 Evaluation**

Disciplinas	Evaluaciones					Asistencias / Ausencias				
	1 P	2 P	3 P	Final	Prom	1 P	2 P	3 P	Final	Total
Manufactura de Prendas I Hernandez Bertha	9	9	10	10	9.5	5/0	5/0	3/1	4/0	17/1
Corte Industrial II	9.5	9.5	9.5	10	9.8					

Système de qualifications

Un système de news permet aux professeurs et aux administrateurs de communiquer aux utilisateurs des groupes de leur choix. Une case à cocher permet d'envoyer le communiqué par email à tous les utilisateurs concernés. Le système de document leurs permet quand à lui de mettre des documents en ligne pour leurs cours. Ceux-ci doivent obligatoirement être au format PDF afin de permettre aux élèves de les consulter sans à avoir à disposer de l'application avec laquelle le document a été créé.



Systeme de documents

## 5. Mise à test parallèle

J'ai mis le nouvel intranet en test sur un domaine séparé *intranet.iesmoda.edu.mx*, tout en laissant l'ancien intranet en fonctionnement (*www.iesmoda.edu.mx/intranet*).

La période de test de l'intranet – de seulement dix jours –, m'a permis de régler de nombreux bugs et de tester l'intranet auprès de plusieurs utilisateurs.

## 6. Ajout de fonctions, transition opérationnelle

Le temps venant à manquer à la fin du stage, j'ai lancé l'intranet sans tarder en mettant une page de redirection à l'adresse de l'ancien intranet et en plaçant ce dernier sur une sur adresse privé, en cas de besoin pour l'administration.

Quelques fonctions supplémentaire on ensuite été ajoutées à chaud comme une module obligeant les utilisateurs à accepter la chartre d'utilisation de l'intranet avant de pouvoir l'utiliser. De nombreuses modifications cosmétiques ont par ailleurs été réalisées après avoir recueilli les commentaires des utilisateurs.

## **Résultats**

Au final la plus grande partie du temps de travail consacré à l'intranet l'aura été via la refonte de celui-ci et sa sécurisation, bien que la demande initiale fût portée sur des corrections de problème et l'ajout de fonctionnalités.

Je ne regrette pas une seconde le choix que j'ai fait, en effet, en reprenant le code initial, j'aurais certainement jamais réussi à atteindre les objectifs. L'intranet est surtout à présent bien plus sécurisé qu'il n'était auparavant, ce qui était impératif vu la sensibilité des informations impliquées.

La taille totale du code source de l'intranet est passée de 475 kilo-octets à 274 kilo-octets, tout en y ajoutant de nombreuses fonctions. La base de données a été rendu plus cohérente, complète et sécurisé, sa structure est, passé de 21 tables à 14 tables. Ce n'est qu'une partie des nombreux facteurs qui prouvent que la maintenance de l'intranet a été rendu beaucoup plus aisée pour l'avenir, afin que les autres personnes qui travailleront dessus n'aient pas à se poser la question de repartir quasiment à zéro.

L'intranet est par ailleurs devenu plus convivial et facile à utiliser pour ses utilisateurs. Des infos contextuelles ont été placées le long des formulaires, et l'interactivité visuelle a été développée, via la possibilité par exemple de naviguer sur le trombinoscope des étudiants d'un groupe.

La directrice de l'IESModa compte donner une importance croissante à l'intranet avec le temps, en y centralisant de plus en plus des fonctions de l'administration de l'école. Elle souhaiterait encore réduire l'utilisation de papier dans le nouveau bâtiment en permettant aux élèves de marquer directement leur présence en classe à partir d'un ordinateur.

Je pense qu'elle a parfaitement compris qu'une informatique maîtrisée et utilisée intelligemment permet de réduire les coûts et les délais de réactivité, tout en donnant une image positive à son école.

## Bilan de la fin du stage

### Tarik Ansari

Ce stage m'a permis d'assurer un projet informatique de type « métier » (l'intranet), ainsi qu'un projet informatique innovant, le tout dans un contexte multiculturel. La directrice n'a pas hésité à nous donner des responsabilités, et nous avons vraiment touché à un échantillon très riche du métier d'ingénieur.

Nous avons réussi à mener à bien deux gros projets avec un planning très serré (4 mois), ce qui me semble honorable. Pour cela, nous avons dû faire preuve d'adaptabilité et d'organisation. Nous avons également été amenés à gérer une équipe, et c'est finalement peut-être là que nous avons appris le plus.

Nous n'avons pas eu le temps de nous ennuyer lors de notre séjour au Mexique, et c'est sans doute ce qui nous a permis de travailler avec tant de plaisir. Cette expérience m'a vraiment donné le goût de travailler en équipe dans un contexte international.

## **Jean Marty**

Lors de ce stage technique, j'ai abordé des problématiques modernes dans les technologies de développement web. Dans ce cadre, j'ai pu développer ma capacité à proposer des idées originales. En effet, nous avons évolué dans un environnement libre des contraintes d'une grosse entreprise, et avons bénéficié des investissements souhaités. Ainsi nous avons les moyens et la liberté d'entreprendre, conditions propices à l'épanouissement d'un ingénieur. J'ai aussi profité de l'expertise de Tarik dans ces technologies pour enrichir mes compétences en développement web, dont je n'avais jusqu'alors que des notions. Enfin, ce stage a stimulé mon envie d'apprendre et m'a fait comprendre que l'organisation, les prévisions et donc la maîtrise de l'information permettent de prendre les décisions plus vite, en étant toujours tourné vers l'avenir.

D'un point de vu humain, je me suis intégré facilement à l'équipe projet et la communication avec les autres stagiaires m'a ouvert à une autre culture, peut être moins formaté qu'en France. Ainsi, J'ai enrichi mes compétences techniques, et créé des contacts professionnels et humains. Pour tout cela, ce stage restera une expérience unique.