

Des informations sur le système de formation grâce au Web

Robert MAHL

Ecole des mines, Paris

Historique

Vers septembre 1994, l'Ecole des mines fut approchée par le Comité d'Etudes sur la Formation d'Ingénieurs (CEFI) pour créer un service d'information sur le Web concernant les Diplômes d'Etudes Approfondies (DEA) français de l'ingénierie, des formations d'un an qui aboutissent au niveau Bac + 5. Cette première opération fut menée en moins de deux mois, et suivie du chargement d'informations sur d'autres cycles de niveau Bac + 5, les DESS et les Ecoles d'ingénieurs.¹

Nous avons ensuite installé sur le Web une base de données (dénommée CANASTA ou FORMATEL) qui répertorie de façon détaillée 43.000 stages de formation professionnelle, dont 10 % environ sont conventionnés, cette base étant maintenue sous l'égide de l'Etat français et de la Région Ile de France.²

La Direction de l'Evaluation et de la Prospective du Ministère de l'Education Nationale souhaite ensuite créer un service Web en s'appuyant sur l'Ecole des mines. Des informations très complètes et illustrées sur l'état du système scolaire français et la géographie de l'Ecole furent installées.³

En liaison avec la Direction Générale des Enseignements Supérieurs, nous avons transporté sur le Web le contenu d'un CD-ROM concernant le système de formation français.⁴

Par ailleurs, l'Ecole des mines héberge un service d'aiguillage vers des services administratifs sur le Web, désigné par "Adminet".⁵

Nous sommes en train de porter la base sur les Ecoles et sur les Mastères de la Conférence des Grandes Ecoles, organisme associatif français qui regroupe 165 écoles de niveau Bac + 5 (environ 130 écoles d'ingénieurs et 35 écoles de commerce), et qui gère les "mastères".⁶

¹ URL : <http://cri.ensmp.fr/cefi/>

² URL : <http://cri.ensmp.fr/stages/>

³ URL : <http://cri.ensmp.fr/dep>

⁴ URL : <http://cri.ensmp.fr/mesr/>

⁵ URL : <http://cri.ensmp.fr/~scherer/adminet/>

⁶ Curieusement, en France la quasi-totalité des diplômes d'enseignement supérieur, même ceux attribués par un organisme d'enseignement privé, ont un statut de diplôme public géré par un Ministère. Les "mastères" dérogent à cette règle puisqu'ils sont gérés par un organisme privé associatif, même lorsqu'il sont attribués par un organisme de formation public !

Nous hébergeons aussi quelques informations pour le compte de la Fédération Européenne des Associations Nationales d'Ingénieurs (FEANI).⁷

L'origine des données, et la problématique de la base de données à accès gratuit

Le CEFI édite depuis fort longtemps des catalogues des formations supérieures : il recense environ 250 DEA scientifiques ou techniques (plus de 500 dans la prochaine édition, sur 1100 formations de DEA en France), environ 200 DESS et 250 écoles d'ingénieurs. Ces guides ou catalogues sont diffusés par le CEFI à des prix qui vont d'environ 100 F pour un guide "grand public" à plusieurs milliers de Francs pour un catalogue professionnel. Le CEFI diffuse aussi des disquettes pour MAC au format Hypercard, et envisage la diffusion de CD-ROM sur l'enseignement supérieur français (de tels CD-ROMs se diffusent déjà très couramment aux Etats-Unis).

Or le système actuel du Web ne permet pas aux serveurs de demander une rémunération de l'information. Il était donc bien légitime que le CEFI puisse craindre que la présence d'un nouveau média d'informations gratuites ne supprime les médias payants, et que de ce fait les revenus du CEFI diminuent. Or plusieurs considérations ont conduit les responsables du CEFI à réaliser néanmoins l'expérience, et probablement à la pérenniser :

1) En 1994, l'ambassade de France aux Etats-Unis s'inquiétait du manque d'informations sur le système scolaire français sur le réseau Internet. A la suite d'une concertation entre la Direction Générale des Enseignements Supérieurs et le Ministère des Affaires Etrangères, il fut alors décidé de réaliser diverses expériences d'information sur le Web. En particulier, un service universitaire de Lyon fut chargé de rassembler des informations sur les Ecoles d'ingénieurs et de les mettre sur disquettes et sur Web, et le ministère de l'Education Nationale sous-traita à une Société de Services la création d'un serveur Web avec l'étiquette ministérielle qui comporte notamment la liste des DEA.⁸ Toutefois, ces informations étaient ne comportent aucune description des programmes des formations concernées. Le CEFI, qui possédait des informations bien plus complètes dans le domaine des formations pour ingénieurs, fut donc encouragé de les mettre sur Internet.

2) La disponibilité d'informations sur le Web peut servir de catalyseur pour inciter le public à acheter les guides complets. Le CEFI a donc souhaité que seulement une partie des informations soit installée. En particulier, les informations nominatives sont généralement occultées, afin d'être en conformité avec les avis de la Commission Nationale Informatique et Libertés (CNIL). Comme l'utilisateur final sait que l'édition papier est plus complète que l'édition électronique, mais comme il ne connaît pas précisément la "différence" entre les deux, il est souvent tenté d'acheter l'édition papier.

3) Une question fréquente des fournisseurs d'information comme le CEFI concerne le coût, pour l'utilisateur final du réseau Internet, d'imprimer l'ensemble des pages d'information qui sont sur le serveur. A l'évidence, si ce coût est inférieur à l'acquisition du "catalogue papier", et comme le délai est moindre, l'utilisateur final privilégiera la solution réseau. Or, force est de constater que le coût de fabrication du guide papier à partir d'Internet est encore

⁷ URL : <http://cri.ensmp.fr/feani/>

⁸ URL : <http://dgrt.mesr.fr/> et <http://dgrt.mesr.fr/SIF3/dspt.html>

élevé ; il faut en effet tenir compte du temps de configuration de l'automate de récupération automatique des pages du Web, des délais de reprise en cas de problèmes réseau, du coût de l'impression laser.

La problématique de la base CANASTA est différente. Cette base fut créée il y a plus de 5 ans pour informer les jeunes chômeurs sur les stages de formation "conventionnés" disponibles. La maintenance de la base a été confiée par l'Etat et la Région Ile de France à une société privée, qui a étendu le domaine de saisie à l'ensemble des stages offerts par des organismes privés ou publics, conventionnés ou non. Depuis 5 ans déjà, cette base est disponible au grand public par Minitel sous le code 3617 PLPA (coût : environ 120 F/heure) et aux professionnels de la formation sous le code 3614 avec mot de passe (accès gratuit). Il pouvait sembler que le Web gratuit ferait concurrence au Minitel payant. Pourtant, c'est avec une rare unanimité que les responsables de la base, Pouvoirs publics ou société privée, ont souhaité mener l'expérimentation sur le Web. L'une des raisons est que les recettes du 3617 ne couvrent qu'environ 20 % du coût d'entretien de la base. Une autre raison est certainement la recherche de notoriété des gestionnaires pour leur produit. Seul le prestataire informatique du serveur Minitel a manifesté une réticence, car il craignait de voir son client lui échapper.

Dans le cas des données provenant de ministères (DEP et DGES), la gratuité est de rigueur. Donc l'implantation des données sur un serveur Web pourrait ne sembler poser que des problèmes techniques. De fait, nous avons rencontré une grande ouverture d'esprit de la part des fonctionnaires responsables. Toutefois, les ministères sont très sourcilleux sur une certaine "propriété intellectuelle" de leurs données, et souhaitent contrôler très étroitement leur mise sur le Web, en y imprégnant leur image. D'autre part, certaines données jouissent de statuts particuliers, comme par exemple les textes législatifs et réglementaires : tout le monde a le droit de "recopier" le Journal Officiel, mais il n'est pas clair juridiquement si sa recopie systématique à partir d'une base informatisée en vue de fabriquer des pages Web nécessite ou non une autorisation préalable (difficile à obtenir en pratique). Nous avons ainsi cherché en vain à obtenir de services publics français la fourniture de fichiers informatisés existant, qui établissent la correspondance entre différents systèmes de nomenclature des actions de formation (FORMACODE, code ROME, code de la nomenclature interministérielle de 1994).

Une mention toute particulière concerne ADMINET. C'est l'oeuvre personnelle d'un haut fonctionnaire, M. Christian Scherer, qui maintient lui-même des batteries de pointeurs vers tous les nouveaux services publics qui naissent sur le Web. En outre, pour les organismes publics qui ne sont pas encore présents sur le Web (comme l'INSEE, La Poste, etc.), M. Scherer trouve des pointeurs vers des pages Web privées relatives à ces services. ADMINET jouit d'un taux de consultation considérable, puisque environ 10.000 pages sont regardées dans les jours de pointe, ce qui le classe parmi les dix services Web les plus examinés. ADMINET a été déterminant pour faire connaître nos produits Web sur la formation.

L'ergonomie de consultation

On accède à l'information utile de la plupart des bases selon différents principes :

1) Pour les "petites" bases, il y a des pages d'index : par région géographique, par ville, par type d'organismes, par domaine de formation, voire par tutelles des écoles d'ingénieurs ou même par mode d'admission (niveau, concours ou examens).

2) On peut aussi taper un mot (ou une fraction de mot) dans une zone de saisie, ce qui crée dynamiquement une page d'index si ce mot est présent dans un dictionnaire (qui a été généré automatiquement au préalable).

3) Dans certains cas (DEA, DESS, Ecoles), une carte cliquable est disponible.

4) La base sur les stages de formation est traditionnellement accessible sur Minitel par le choix de un ou plusieurs critères parmi une dizaine de critères disponibles : domaine, conventionnement, département géographique, diplôme en fin de stage, niveau, ... Nous avons reproduit ce mode d'interrogation sur le Web en utilisant des formulaires. Evidemment, la saisie d'une question qui nécessite typiquement de consulter une dizaine de pages Minitel nécessite une seule page Web, ce qui est bien plus satisfaisant. Néanmoins, nous avons remarqué que les utilisateurs naviguent aussi souvent par saisie de mots que par formulaires ; de fait, la base incorpore une arborescence des domaines de formation à 3 niveaux intermédiaires, mais le suivi des trois étapes de choix tel qu'il se fait habituellement sur Minitel ne garantit pas que l'on arrive à l'information utile : si quelqu'un cherche un stage sur le multimédia, doit-il suivre le chemin "Communication - Télévision - Traitement de l'image", ou bien le chemin "Informatique - Techniques de pointe - Image et son", ou encore un autre ?

5) Dans le cadre de l'interrogation de la base des DEA, nous disposons d'une dizaine de mots-clés décrivant chaque DEA, fournis par le responsable de chaque DEA mais n'appartenant pas à un dictionnaire préétabli de mots-clés. Ceci nous a permis d'établir un dispositif original de navigation par proximité sur les mots-clés. Par exemple, on observe le DEA intitulé "Science des aliments" ; celui-ci comporte notamment le mot-clé cliquable "Biochimie des Aliments". En le sélectionnant, l'utilisateur non seulement tombe sur le DEA "Biotechnologies et Industries Alimentaires" qui a par hasard le même mot-clé, mais on lui donne également une liste de mots-clés voisins parmi lequel "Physiologie et biochimie végétales". En cliquant sur ce dernier mot-clé, on arrive au DEA "Bases de la Production Végétale", et ainsi de suite.

Si la présentation des pages d'information textuelles elles-mêmes n'a pas posé de vraie difficulté, il en va différemment des images et des pages d'accueil des services.

Il y a un an, lorsque nous avons démarré ces travaux, les clients du Web étaient habituellement équipés de MOSAIC qui ne savait pas afficher le format JPEG. De ce fait, le format GIF a été utilisé pour la conversion du CD-ROM de l'Enseignement supérieur, pénalisant les photos d'un facteur 2,5 en moyenne par rapport à JPEG, et causant des délais de transmission inacceptables sur le réseau. Comme désormais JPEG est disponible grâce à l'outil NETSCAPE, nous n'utilisons plus désormais que des images JPEG pour les photos (GIF pour les graphiques) dont la taille doit si possible être inférieure à 5 K Octets. D'autre part, nous avons réalisé avec la DEP une expérience malheureuse de scanning de quelques centaines de graphiques, mais la qualité des images sur les PC de basse définition des clients était tellement déplorable que le Ministère a préféré les refabriquer directement grâce à un tableur, en limitant à 2 couleurs pour avoir une taille réduite. A cet égard, nous attendons avec impatience la technologie "JAVA" de SUN qui permettra de générer les images chez le client.

Les fournisseurs étaient généralement très exigeants sur les pages d'accueil de leurs services, et ont tenu à les réaliser eux-mêmes. La dualité des langages français-anglais nécessitait généralement plusieurs itérations pour arriver à une présentation satisfaisante. Nous avons conseillé aux fournisseurs d'information d'ouvrir des boîtes aux lettres sur Internet, afin

que leurs pages d'accueil fournissent un lien direct de messagerie ou voie de retour permettant aux utilisateurs de faire des commentaires.

Le service de la DEP permet notamment le téléchargement d'une disquette de type PC qui contient des données statistiques. Ce type de disquette est traditionnellement expédié chaque année à tous les chefs d'établissements scolaires français. Désormais, le grand public a également accès par Internet à ces données.

La transformation des données en pages Web

Les données peuvent se trouver initialement :

- soit sous forme d'une base de données organisée (CANASTA/FORMATEL est ainsi composée de 5 tables principales totalisant 30 Méga Octets d'information utile),

- soit sous forme de traitement de textes : les ouvrages du CEFI étaient en partie sous FrameMaker, d'autres documents furent récupérés sous Word de Microsoft ; dans ce cas il fallait les transformer par des filtres en fichiers balisés afin de constituer automatiquement une vraie base de données,

- soit sous forme papier ou même sonore ; ainsi, les 700 images du CD-ROM de la DGES ont pu être récupérées automatiquement, mais comme il n'était pas raisonnable de faire entendre les bandes sonores par la Web pour des raisons pratiques de débit, il a fallu saisir les paroles en texte ASCII puis créer un véritable scénario, ce qui nous a amené incidemment à développer un petit langage qui permet de formaliser l'expression de l'enchaînement des pages de texte et d'images.

Lorsque la base de données est finalement disponible, la transformation en fichiers HTML (format des documents sur le Web) est totalement automatisée, y compris la création des liens hypertextuels. L'ensemble de ces transformations est décrit grâce au langage PERL ou au langage AWK : ce sont des langages interprétés, extensions du langage C ; de plus, PERL 5.0 est orienté objet. Nous avons aussi testé l'utilisation du TOPIC DEVELOPMENT KIT de VERITY INC. pour la recherche documentaire.

L'automatisation des transformations permet de rejeter toute erreur factuelle sur le producteur de la base, et permet un rechargement rapide d'une base. D'autre part, nous avons le choix entre le procédé habituel dans les serveurs Minitel français d'adosser le service à des requêtes SQL transmises à une base de données et la création statique de pages HTML complètes, y compris pour les index : nous avons choisi généralement cette dernière méthode, dont l'inconvénient est de multiplier par un facteur 5 à 10 la taille des bases sur disque, mais qui permet d'accélérer le traitement des requêtes, d'économiser le coût des licences de SGBD relationnels, et par là même de transporter facilement une base sur PC ou MAC pour une consultation locale avec la même ergonomie.

Par ailleurs, nous souhaitons que la description d'une action de formation comporte un pointeur vers le serveur Web de l'organisme (lorsque celui-ci existe). Nous avons donc constitué un fichier qui comporte actuellement quelques centaines d'entrées, avec les URL des universités, écoles, organismes divers, et qui donne aussi la correspondance entre les noms des organismes dans les différentes bases que nous gérons (par exemple, le code "X" dans une base correspond à "POLYTECHNIQUE MASSY" dans une autre, et "EP" dans une troisième. Ceci permet dans certains cas de créer des liens hypertextuels entre différentes bases : ainsi, à partir d'une école de la base des écoles on peut accéder à ses DEA ou DESS dont les pages Web

sont créées indépendamment à partir d'autres bases. Nous n'avons toutefois pas établi pour l'instant de lien entre des bases de différents fournisseurs, car l'acceptation de ces liens de la part des fournisseurs d'information s'avère délicate : on ne va pas directement de la page "actions de formation continue de l'Ecole des Ponts" à la page "DEA de l'Ecole des Ponts".

La rançon du succès ?

Les taux de fréquentation de nos différents services⁹ augmentent exponentiellement, de même que le trafic d'Internet. Les bases et documents du CEFI sont consultés pour un peu plus de la moitié par un public français, pour le reste par des étrangers dont la plupart nord-américains. La base sur les stages et les documents de la DEP ont un public essentiellement français, quoique les nombreuses pages et graphiques sur l'état de l'école soient disponibles en français et en anglais.

L'augmentation du taux de fréquentation a été obtenu sans aucune publicité particulière ni communication à la presse. Deux moteurs de diffusion ont été déterminants :

- ADMINET et l'UREC, qui sont des pages d'hyperliens mises à jour manuellement,
- des moteurs d'indexation automatique anglo-saxons, aussi appelés "robots" ou "knowbots" ou "surfers", parmi lesquels InfoSeek, Lycos ou WebCrawler.

Par ailleurs, il se crée progressivement une attitude dans la profession des services Web d'échanger les pointeurs : en échange du service que vous rend autrui de vous référencer par un pointeur, ce dernier espère de votre part (même s'il ne vous le demande pas explicitement) que vous établirez un pointeur en retour vers le sien. Ce genre de mécanismes joue en faveur de la fréquentation tant d'ADMINET que des bases du CEFI, car bien des Ecoles ou Universités référencées établissent des pointeurs en retour. Ceci pose d'ailleurs un problème technique, car beaucoup de pages ont des noms (URL) générés automatiquement par les filtres de création, dont la pérennité n'est pas garantie : pour éviter que des pointeurs ne tombent dans le vide ou au mauvais endroit lors de la régénération des bases, il faudra adopter dans l'avenir des dispositifs de nommage symbolique et surtout les faire connaître aux référenceurs (ce qui pose le problème bien connu que le protocole actuel du Web ne vous permet pas de connaître les URL des pages qui pointent vers une de vos propres pages).

La mise à jour des bases directement à partir du Web est un enjeu majeur pour l'avenir. Il se trouve que la base CANASTA/FORMATEL a une bonne expérience dans ce domaine du fait de ses 5 ans de présence sur Minitel (et 1400 heures par mois d'accès). Les organismes de formation peuvent obtenir un mot de passe qui les autorise à taper gratuitement sur Minitel les informations à introduire dans la base, la validation étant néanmoins faite manuellement par les responsables de la base. Il s'avère que les centres de formation privés sont généralement très disposés à faire cet effort, mais que les centres publics (et notamment les grandes écoles d'ingénieurs, les universités ou l'AFPA) sont généralement moins alertes sur cette forme de

⁹ Voir à ce sujet <http://cri.ensmp.fr/stages/wusage/> , <http://cri.ensmp.fr/cefi/wusage/> , <http://cri.ensmp.fr/dep/wusage/>

Les courbes et chiffres donnés n'incluent pas les consultations "internes" provenant des fournisseurs d'information et de l'Ecole des mines. Ils comprennent par contre les accès des robots de recherche (environ 20 % du trafic total). L'ensemble des accès (mais non compris ceux à ADMINET) représente environ 60.000 pages chargées par environ 3.000 utilisateurs différents en novembre 1995.

communication. Sans vouloir tirer de conclusion générale, on peut supposer que le marketing des services publics puisse encore faire des progrès ! En attendant un tel dispositif de mise à jour directe, la nouvelle campagne de mise à jour de la base des DEA va s'appuyer sur 4 associations d'universitaires par grandes disciplines, qui obtiennent les informations par voie "papier" de leurs adhérents, et les saisissent elles-mêmes.

Beaucoup d'utilisateurs nous envoient des messages. Ceux-ci sont désormais transmis en parallèle au fournisseur de l'information et au responsable du serveur à l'Ecole des mines. On peut distinguer 3 catégories de messages :

- les messages d'étudiants cherchant des renseignements : il s'agit généralement d'étrangers qui posent les questions les plus inattendues comme le niveau des bourses, ou bien comment chercher un logement à Paris. Nous essayons de leur répondre systématiquement, même brièvement, quitte à leur envoyer une adresse utile où ils peuvent se renseigner ;

- les demandes de modification d'informations présentes sur le Web. Ce sont généralement des modifications de détail. Souvent, on nous demande d'introduire un nouveau pointeur ou de modifier un pointeur existant. On ne nous a encore jamais demandé de supprimer une information quelconque, même pas une information nominative ;

- les messages de sympathie et les propositions de collaboration, toujours bienvenus.

Plusieurs organismes (dont la FEANI, Fédération Européenne d'Associations Nationales d'Ingénieurs) réfléchissent à l'opportunité de mettre sur Internet des listes d'ingénieurs ou de diplômés. Différentes tentatives ont déjà réalisées dans ce sens en France. Toutefois, le contrôle de la CNIL sur ce genre de bases est redoutable, ce qui délaie ce genre d'expérience qui pourtant serait techniquement très facile à réaliser.

Avec le temps, bien sûr, l'ambition des fournisseurs d'information augmente. Ainsi, ADMINET se constitue depuis peu une bibliothèque de textes administratifs relatifs à l'éducation (en texte intégral). Nous avons réalisé sur le Web un petit système expert sur les rémunérations des stagiaires de formation, mais le Conseil Régional Ile de France nous a demandé de surseoir à sa mise en service car il n'a (évidemment !) pas la preuve de son bon fonctionnement, et la responsabilité juridique en cas de plaintes n'est pas claire. La base CANASTA/FORMATEL sera enrichie d'ici fin 1996 d'une quantité de textes décrivant des métiers, avec une navigation de proximité d'un métier à un autre ainsi que des métiers aux formations et réciproquement. Nous avons également l'ambition de participer à des projets européens dans ce domaine, ce qui nécessite une bonne segmentation des usagers et des profits qu'ils en tirent (par exemple : organisation de stages d'étudiants, simplification de procédures d'inscription dans les universités, programmes d'échanges gradués ...).