

Dynamiser ses signets

Une des missions des documentalistes étant de repérer et valider l'information sur l'Internet, pourquoi ne pas construire à l'attention de ses utilisateurs une base de données de signets ? Deux approches sont possibles.

Chacun connaît l'importance des signets (raccourcis vers un site web) : nous avons tous une liste de sites plus ou moins organisée que nous conservons soit sur notre bureau, soit dans les favoris ou bookmarks de notre navigateur. Mais pour un organisme documentaire comme une bibliothèque ou un centre de documentation, on ne parle plus de liste, mais bien de collection car les liens vers les ressources Internet font partie intégrante de l'offre documentaire.

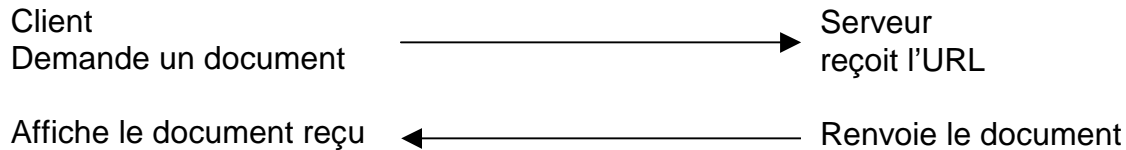
La façon la plus simple de publier une collection de signets reste la création et la mise à jour de quelques pages web statiques (1) à l'aide d'un éditeur HTML (Hypertext markup language). L'observation des portails sur Internet nous livre une façon efficace de présenter les signets. Une page contient les rubriques correspondant à un plan de classement hiérarchique qui servira à organiser les signets. Nous avons donc une page d'accueil qui présente de façon concise les thèmes qui servent de lien vers un ensemble de pages qui contiennent la description des sites sélectionnés.

Une expérience menée à l'Enssib

Cependant, plus le nombre de signets augmente, plus la collection est difficile à maintenir. La vérification de la validité des signets est compliquée ainsi que leur mise à jour. L'utilisation d'une base de données peut s'avérer utile. Plus besoin alors de jongler avec le texte pour ajuster la mise en page. On peut aussi ajouter des informations de gestion aux informations qui servent à la recherche. Par exemple, l'ajout d'un champ date, correspondant à la dernière visite du site, permet de trier les signets par ordre décroissant dans le but de revisiter les sites les plus anciens. Un champ auteur, nom de la personne qui a trouvé le site, devient intéressant dans le cas d'une gestion partagée de la collection. Le bibliothécaire ou le documentaliste est donc confronté au choix d'un outil de gestion ainsi que d'un mode de publication des signets. Cet article résume une expérience menée à l'ENSIB (2) dont l'objectif était de comparer deux méthodes de publication d'une base de données sur un intranet ou Internet. La première, développée à l'aide du langage de programmation PHP (Personal home page tools), met en oeuvre une base de données relationnelle, appelée MySQL (www.mysql.org), et la seconde utilise un moteur d'indexation, en l'occurrence Aurweb de la société Auracom (www.auracom.fr).

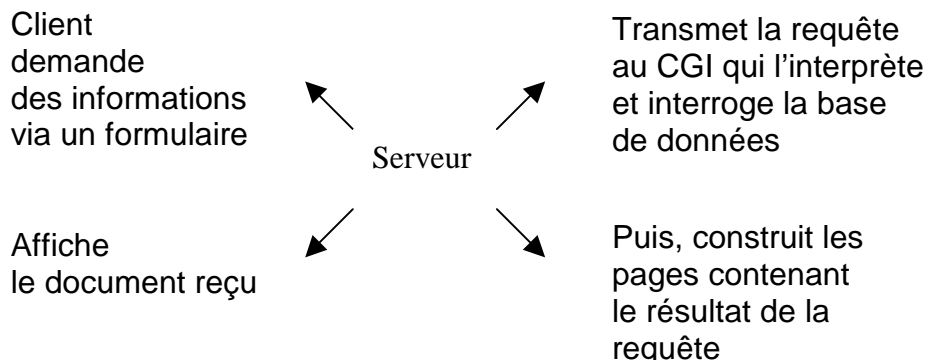
La gestion des signets à l'aide d'une base de données permet généralement d'abandonner la maintenance manuelle des pages web. Un programme d'exportation génère dynamiquement les fichiers HTML (pages web statiques) en insérant le contenu de chaque enregistrement à l'intérieur des balises HTML nécessaires (cf. fig.1 : requêtes statiques et dynamiques).

Requête statique = URL correspondant à un nom de fichier stocké sur le disque du serveur



Exemple : <http://www.enssib.fr/recherche/bibliouc/index.html>

Requête dynamique = URL correspond à l'appel d'un script CGI



Exemple : <http://cdrap.cd-script.fr/aurweb.exe/voir?critA=xml&critB=web>

Ici, le fichier envoyé n'existait pas. Il a été créé à la volée par le CGI à partir du résultat de l'interrogation de la base de données

Fig.1 : requêtes statiques et dynamiques

Consultation par navigation et par interrogation

La possibilité d'offrir l'interrogation en ligne de vos signets n'est pas incompatible avec leur publication sous forme statique. Il est tout à fait possible d'insérer dans votre page d'accueil un petit formulaire: qui effectuera une recherche dans la base de données. De cette façon, le public dispose de deux moyens de consulter votre collection : par navigation et par interrogation.

La publication d'une base de données sur un intranet ou sur Internet repose sur le serveur web de l'organisme. Ce dernier peut gérer plusieurs ressources simultanément (sites ou bases de données). La déclaration d'un alias, c'est à dire un lien entre une URL (3) et un répertoire sur le serveur, permet d'accéder à la ressource. Par exemple, l'URL <http://www.organisme.fr/dbSignets/> pourrait représenter le répertoire d:\siteweb\signets du serveur, dans lequel se trouvent les fichiers nécessaires à l'interrogation de votre collection de signets. L'opération de création d'alias est réalisée par l'administrateur du réseau de votre entreprise.

Pour offrir l'accès à une base, les trois éléments suivants doivent être installés sur le serveur web :

- une base de données (ou des fichiers structurés, selon la technologie utilisée) ;
- un formulaire de recherche inséré dans une page web (c'est cette page qui sera envoyée à l'internaute qui voudra interroger votre base) ;
- un script CGI qui interprète les critères saisis dans le formulaire, interroge la base de données et construit la page à renvoyer avec le résultat de l'interrogation (un CGI est un petit programme stocké sur le serveur et qui est invoqué par l'attribut "action" du formulaire):

Fig.2 : création de la base de données

Champ

Nom	Nom du site web
Adresse	Adresse du site (URL)
Thème	Thème du plan de classement
Description	Commentaire sur la ressource
Vdate	Dernière visite du site

Exemple d'un enregistrement

Nom	CRAC (Centre de recherche et d'actions culturelles de Valence)
Adresse	http://www.crac.asso.fr
Thème	Culture
Description	Association valentinoise dont l'objectif principal est de proposer des animations culturelles sur plusieurs aspects de la culture. Elle gère une salle de cinéma dont la programmation comprend des films d'art et d'essai. S'intéresse au multimédia.
Vdate	05/02/2002

I - création de la base de signets

La première méthode retenue pour notre expérience s'appuie sur une base de données de type SQL (4). Nous avons d'abord récupéré le package EasyPHP (5), qui contient un serveur web, Apache, un serveur de bases de données SQL, MySQL, et un langage de script, PHP. Ce package facilite l'installation d'une configuration complète car il permet de s'affranchir des paramétrages de chacun de ces éléments. L'installation d'EasyPHP sur un poste en local est immédiate, le raccourci vient s'insérer dans le menu Démarrer de Windows et un simple clic suffit pour lancer le serveur Apache. La page d'accueil qui s'affiche permet de définir l'emplacement de votre site web et donne accès à une interface conviviale de gestion de la base de données MySQL : PHPMyAdmin.

Nous commençons par créer une première table contenant les cinq champs d'un enregistrement : nom, adresse, thème, description, date de validation (cf. fig.2). Nous gérons les droits d'accès à partir d'une autre table contenant le login et le mot de passe de chacune des personnes autorisées à modifier la base.

2 -création d'un formulaire

C'est à travers le formulaire que l'utilisateur transmet sa requête à la base de données. Notre formulaire offre deux possibilités de recherche: par mot sur trois champs combinés (nom, thème, et description) et par thème à partir d'une sélection dans une liste. Pour créer un formulaire de recherche, trois éléments du langage HTML sont importants (6) : la balise <Form>, la zone <Input> et le bouton <SUBMIT>. Voici le code de notre formulaire, commun aux deux solutions (cf. fig. 3).

L'attribut action de la balise <Form> a comme valeur le nom du programme ou du script CGI, les trois autres lignes (Input de type text) décrivent les champs avec les libellés en gras (), suivis des noms des champs qui contiendront les informations saisies par l'utilisateur. Puis, avant la fermeture de la balise <Form>, dans une zone Input de type submit (un bouton) qui déclenche l'envoi de la requête. Par exemple, si l'utilisateur a saisi « bpi » dans la fenêtre correspondant au champ Nom, l'URL transmise au serveur sera : http://www?serveur.fr/? Nom= bpi. En plus du formulaire de recherche, on peut en créer d'autres destinés à la gestion de la base. Plusieurs personnes peuvent ainsi participer au développement de la collection de signets. L'accès à ces derniers est contrôlé par mot de passe.

```
<FORM ACTION="/cgi-bin/program.exe">
<B>Mot-clé</B><INPUT TYPE="text" NAME="nom">
<B> Theme</B><SELECT name="Theme">
<OPTION selected value=""></OPTION>
<OPTION value="Thème1">Thème1</OPTION>
<OPTION value="Thème2">Thème2</OPTION>
<OPTION value="Theme3">Thème3</OPTION>
</SELECT>
<INPUT TYPE="SUBMIT" VALUE="Chercher"> </FORM>
```

3 - Gérer les requêtes entre l'utilisateur et la base de données

Une fois avoir conçu la page HTML qui comprend le formulaire de recherche, il est nécessaire de créer les autres pages qui contiendront les scripts CGI (écrits ici en PHP). D'abord, un script qui établit la connexion au serveur et l'ouverture de la base de données ; un autre pour construire la requête avec les termes saisis dans le formulaire par l'utilisateur ; enfin, celui qui génère la page contenant le résultat de la requête. Le résultat apparaît sous la forme d'une liste avec libellé et contenu des champs (7).

Le langage de programmation PHP est simple, mais suppose que vous ayez l'habitude d'utiliser un langage de commande (8). L'écriture de scripts permettant de faciliter la recherche (par exemple la réduction des pluriels) ou de combiner des mots à l'aide des opérateurs booléens demande un certain savoir-faire. La connaissance du langage de requête SQL est essentielle.

Voici un exemple d'une requête SQL correspondant à la saisie du terme «cinéma» dans le champ Thème du formulaire :

```
SELECT * FROM dbSignets
WHERE Theme = "cinema"
ORDER BY Nom
```

Qui veut dire: sélectionne tous les champs du fichier dbSignets où le mot cinéma apparaît dans le champ Thème et tri les résultats par le champ Nom. Cela paraît facile, mais lorsque la requête comprend plusieurs termes combinés avec des opérateurs booléens, l'écriture du script devient compliquée.

Séparer la gestion de la publication

La seconde méthode examinée consiste à séparer la gestion de la base de données de sa publication (9). Les bases de données relationnelles sont orientées vers la gestion de l'information. Leur publication sur Internet nécessite des développements importants.

Un moteur d'indexation est un outil qui indexe des documents à caractère textuel, comme des pages HTML ou XML (eXtensible markup language) ou des fichiers au format texte, structurés à l'aide de délimiteurs (caractères séparateurs d'enregistrements, de champs et de sous-champs) et qui vous permet de concevoir une interface de consultation adaptée à votre public. Un moteur comprend deux applications : un module d'indexation et un de recherche. L'interface d'indexation vous permet de spécifier les fichiers à traiter et la façon de les indexer. Nous avons utilisé le moteur Aurweb (10). La première étape est l'exportation des données sous forme d'un fichier texte délimité. Nous avons nommé ce fichier signets.txt. Puis il faut saisir quatre fichiers de paramètre: signets.dsc, signets.flt, signets.equ et signets.vid. Le fichier signets.dsc contient la structure de votre base; signets.flt, la définition des règles d'indexation (par exemple, ne pas distinguer le singulier du pluriel) ; signets.equ, une liste des équivalences entre les mots ou les expressions (PME = petites et moyennes entreprises) ; et signets.vid, une liste de mots vides (le, la, les, de, du, au, etc.). Aurweb indexe le fichier texte en tenant compte de tous ces paramètres. Le résultat de l'indexation est un fichier optimisé, accompagné d'un index, le tout faisant à peine plus d'octets que votre fichier de départ. Aurweb ne nécessite aucune installation. Dans le répertoire du serveur reconnu par l'alias, vous n'avez qu'à copier les éléments suivants :

- le fichier html qui comprend le formulaire de recherche (index.html) ;
- un sous-répertoire \cgi-bin, qui comprend Aurweb ;
- un sous-répertoire \data, qui comprend les fichiers indexés et de paramètres ;
- un sous-répertoire \html, qui comprend des pages modèles.

En plus de la page index.html, un certain nombre de pages, appelées modèles, seront appelés en fonction du nombre de réponses à la requête. Ainsi s'il n'y a aucune réponse à la question posée par l'utilisateur, la pageO.html est affichée : « Désolé, il n'y a pas réponse à votre requête. » S'il y a une seule réponse, la page1.html, affiche l'enregistrement trouvé. Et s'il y a plusieurs réponses, on affiche la pageN.html dans laquelle on a placé un tableau prêt à recevoir les enregistrements. Les réponses peuvent être affichées par groupe de n enregistrements (généralement 10 ou 20).

Pas tes mêmes moyens

On le comprend, les deux méthodes étudiées ne mettent pas en oeuvre les mêmes moyens. MySQL utilise un langage d'interrogation normalisé. Cette solution présente également l'avantage de cumuler les fonctions de gestion et de consultation.

Cependant, même s'il est vrai qu'un grand nombre de bibliothécaires et de documentalistes commencent à s'approprier les langages de programmation, particulièrement dans le domaine de la gestion de bases de données, la maîtrise du langage SQL ne s'acquiert pas en un jour. Elle est plutôt du ressort des informaticiens.

En revanche, des outils comme Aurweb vous permettent de développer des services d'interrogation de base de données en ligne en vous affranchissant de la connaissance technique d'un langage de requête. Ces outils ont été conçus pour répondre à des besoins documentaires. Par ailleurs, la séparation des fonctions de gestion et de consultation peut parfois résoudre des problèmes de sécurité des données.

Le choix de la solution dépend donc des ressources que vous possédez. Si vous avez un service informatique qui possède une expertise en SQL, l'achat d'une licence MySQL pour Windows est très abordable, sinon nous vous recommandons l'achat d'une licence Aurweb qui vous coûtera plus cher, mais vous fera économiser du temps.

* Professeur associé à l'Enssib
* * Bibliothèque universitaire d'Evry- Val d'Essonne

(1) L'expression "pages statiques" signifie que les pages web qui contiennent les signets sont stockées telles quelles sur le serveur de votre organisme.

(2) Ecole nationale supérieure des sciences de l'information et des bibliothèques : www.enssib.fr

(3) L'URL, pour Uniform resource locator; est l'adresse d'un document sur Internet. Il comprend trois parties : le protocole, le nom du serveur et la localisation du .fichier sur le serveur:

(4) SQL (Standard query language) peut être traduit par langage normalisé de requête. Il s'agit d'un langage spécialisé dans la manipulation des données pour les bases de données relationnelles.

(5) EasyPHP (www.easyphp.org) a été réalisé par Laurent Abbal, Emmanuel Faivre et Thierry Murail.

(6) Lisa Schmeiset: Maintenance et mise à niveau d'un site web. Paris : Sybex, 1999, p.355.

(7) Une démonstration de cette méthode est disponible en ligne : ginb.free.fr

(8) J-C. Houpiet et N. Morin, conservateurs à la bibliothèque de médecine de l'université Henri Poincaré de Nancy, ont diffusé une méthode pour réaliser un annuaire avec EasyPhHP : scd.uhp-nancy.fr/SCD_UHP/SCD_MEDIPRODODFREE_scripts/phannuaire.htm

(9) Une démonstration de cette méthode est disponible en ligne : cdrap.cd-script.fr/index_signets.html (10) Disponible à partir de 6100 (8 accès simultanés).