

Utilisation et traitement des données saisies sous EndNote avec des scripts PHP/MySQL

Sabine Julien¹ et Jacques Maratray¹

Résumé : Cet article présente la façon d'exploiter ou de traiter des données de références bibliographiques saisies ou importées dans EndNote en dehors de leur logiciel natif en utilisant des scripts PHP/MySQL. Deux raisons nous ont amenés à utiliser ce traitement : soit EndNote ne pouvait pas faire des opérations particulières, soit il le pouvait mais de façon trop fastidieuse. Dans ce but nous exposerons dans un premier temps la façon dont les données sont organisées par EndNote, puis la méthode pour les exploiter et enfin nous donnerons quelques exemples d'utilisation.

Mots clés : EndNote, bibliographie, base de données, MySQL, PHP

Introduction

EndNote est un logiciel très répandu à l'Inra. MySQL est un système de base de données open source également très utilisé à l'Inra depuis les serveurs de centre jusqu'à une utilisation individuelle via EasyPHP, qui permet par exemple de générer des pages web contenant les données existantes dans la base et de gérer les données de la base, tout cela en écrivant des scripts en langage PHP. Or nous allons voir que les données d'une base EndNote sont en fait formatées au format MySQL ce qui va permettre de les exploiter en dehors de leur logiciel.

Les noms des champs de la base cités dans cet article sont ceux qui apparaissent dans la base de données et peuvent avoir des noms différents dans EndNote ainsi Record Number dans EndNote est en fait id dans la base.

1. Les données dans EndNote

Une base EndNote, à partir de la version 8, se compose d'un fichier .enl qui est d'un point de vue informatique un fichier compressé au format zippé (.zip). Lorsque l'on ouvre le fichier .enl, par exemple biblio.enl, EndNote le décompresse et organise les données dans un sous répertoire nommé du nom du fichier .enl suivi de .Data, dans notre exemple biblio.Data. Dans ce sous répertoire se trouve toujours un répertoire nommé rdb qui contient toutes les références de la base.

Le répertoire rdb correspond aux fichiers d'une base de données MySQL (**figure 1**) et contient donc les données de la base organisées selon ce format ce qui va permettre leur exploitation et leur traitement.

Nous verrons plus loin que le répertoire .Data, peut recevoir d'autres sous-répertoires comme celui nommé PDF contenant les fichiers liés à la base (à partir de la version X d'EndNote).

¹ UMR1324 CSGA, Centre des sciences pour le goût et l'alimentation - INRA F- 21065 Dijon cedex

☎ 03 80 69 35 31 ✉ Sabine.Julien@dijon.inra.fr (Documentation) –

☎ 03 80 69 30 87 ✉ Jacques.Maratray@dijon.inra.fr (Informatique)

csort.frm	9 Ko	Fichier FRM
csort.MYD	27 Ko	Fichier MYD
csort.MYI	2 Ko	Fichier MYI
db.opt	1 Ko	Fichier OPT
jterms.frm	9 Ko	Fichier FRM
jterms.MYD	0 Ko	Fichier MYD
jterms.MYI	1 Ko	Fichier MYI
misc.frm	9 Ko	Fichier FRM
misc.MYD	11 Ko	Fichier MYD
misc.MYI	3 Ko	Fichier MYI
refs.frm	11 Ko	Fichier FRM
refs.MYD	2 740 Ko	Fichier MYD
refs.MYI	29 Ko	Fichier MYI
refs_ext.frm	9 Ko	Fichier FRM
refs_ext.MYD	0 Ko	Fichier MYD
refs_ext.MYI	1 Ko	Fichier MYI
terms.frm	9 Ko	Fichier FRM
terms.MYD	0 Ko	Fichier MYD
terms.MYI	1 Ko	Fichier MYI

Figure 1 : contenu d'un répertoire rdb. Les références sont contenues dans les fichiers refs.

2. L'exploitation des données

Comme le répertoire rdb est une base de données on peut donc le copier dans le répertoire data d'une installation d'EasyPHP. En utilisant phpMyAdmin, on peut visualiser le contenu de la base (**figure 2**).

Table	Action	Enregistrements	Type	Interclassement	Taille	Perte
<input type="checkbox"/> csort		0	MyISAM	utf8_general_ci	1,0 Ko	-
<input type="checkbox"/> jterms		441	MyISAM	utf8_general_ci	60,3 Ko	-
<input type="checkbox"/> misc		35	MyISAM	utf8_general_ci	24,8 Ko	-
<input type="checkbox"/> refs		2 209	MyISAM	utf8_general_ci	2,7 Mo	-
<input type="checkbox"/> refs_ext		0	MyISAM	utf8_general_ci	1,0 Ko	-
<input type="checkbox"/> reftype		0	HEAP	utf8_general_ci	0 Octets	-
<input type="checkbox"/> terms		2 281	MyISAM	utf8_general_ci	132,8 Ko	-
7 table(s)	Somme	4 966	--	utf8_general_ci	2,9 Mo	0 Octets

Figure 2 : la base de données rdb vue dans phpMyAdmin

Ces données pouvant être traitées comme une base MySQL, il suffit d'écrire des scripts en PHP permettant de réaliser les traitements désirés, soit pour exploiter les données, soit pour les modifier.

Si l'on modifie les données, il faudra ensuite recopier le répertoire rdb dans le répertoire Data de la base d'origine. Il faut ensuite ouvrir la base de données à l'aide d'EndNote puis faire une petite modification anodine (par exemple enlever une lettre et la remettre), et fermer la base dans EndNote. C'est uniquement lors de la fermeture, en ayant fait au préalable une modification qu'EndNote modifiera le fichier .enl. Si aucune modification n'est faite dans

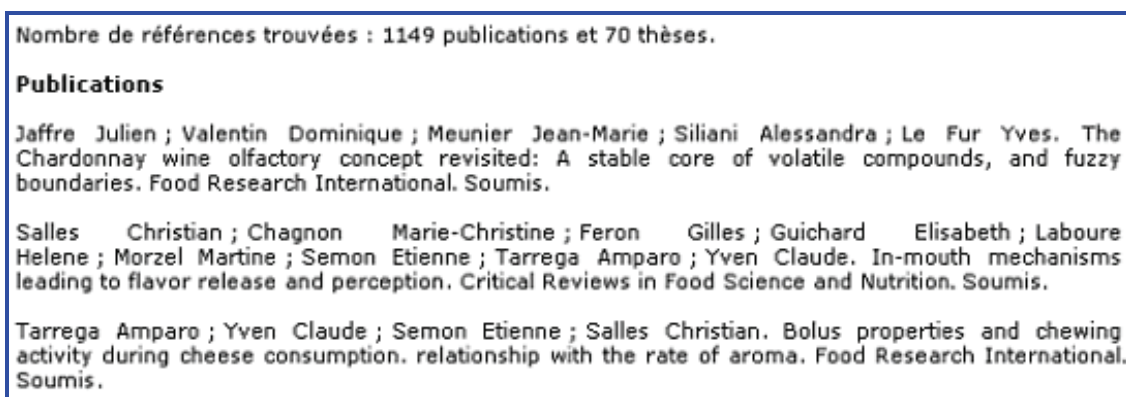
EndNote, le logiciel se contente d'utiliser les données contenues dans le répertoire .Data mais ne refait pas le fichier à la fermeture de celui-ci.

Il est également à noter que les noms des champs de la base MySQL ne correspondent pas toujours aux noms des champs d'EndNote, ceux-ci pouvant être renommés dans l'application alors que les champs de la base sont fixes.

3. Exemples d'utilisation

3.1 Utilisation des données dans un site web

Le premier exemple est une simple récupération des données en vue de leur affichage. Lors de l'élaboration d'un site web dans lequel on veut, par exemple, afficher la liste des publications du laboratoire, il est facile de faire un script PHP d'affichage en utilisant les données de la base (**figure 3**).



Nombre de références trouvées : 1149 publications et 70 thèses.

Publications

Jaffre Julien ; Valentin Dominique ; Meunier Jean-Marie ; Siliani Alessandra ; Le Fur Yves. The Chardonnay wine olfactory concept revisited: A stable core of volatile compounds, and fuzzy boundaries. Food Research International. Soumis.

Salles Christian ; Chagnon Marie-Christine ; Feron Gilles ; Guichard Elisabeth ; Laboure Helene ; Morzel Martine ; Semon Etienne ; Tarrega Amparo ; Yven Claude. In-mouth mechanisms leading to flavor release and perception. Critical Reviews in Food Science and Nutrition. Soumis.

Tarrega Amparo ; Yven Claude ; Semon Etienne ; Salles Christian. Bolus properties and chewing activity during cheese consumption. relationship with the rate of aroma. Food Research International. Soumis.

Figure 3 : affichage des données dans un site web (Réalisé par M. Genty)

On peut également réaliser une recherche sur cette base et afficher les résultats à l'aide du même script.

La mise à jour des publications se fera uniquement en transférant les données de la nouvelle base (répertoire rdb) vers le répertoire data de MySQL. On voit immédiatement l'intérêt car cela élimine toute nouvelle saisie forcément longue et sujette à erreurs.

3.2 Vérification des liens URL et link to pdf dans une base

Une base EndNote a fréquemment des liens, soit vers des sites web, soit vers des fichiers pdf qui sont les tirés à part. Au fil du temps il peut être intéressant de savoir si ces liens sont toujours valides. Pour être accessibles à tous, dans notre laboratoire, les bases ainsi que les tirés à part sont situés sur un serveur web. Compte tenu du nombre d'enregistrements de la base (environ 15 000 références) et du nombre de tirés à part (environ 12 000), la vérification des liens était impossible avec EndNote.

En utilisant un script PHP il est facile de lire les liens enregistrés dans la base et à l'aide de la commande fopen du php -qui fonctionne sur les liens en http- on vérifie rapidement l'intégrité de ces liens et l'on peut générer une page web donnant les liens morts constatés ou même les supprimer directement dans la base.

Dans notre cas, nous avons également fait l'inverse en vérifiant que tous les fichiers présents avaient bien un lien dans la base.

En théorie, il est possible de vérifier les liens web vers les sites externes (ex : <http://abstracts.iovs.org/cgi/content/abstract/49/5/...>) mais comme on ne maîtrise pas la disponibilité des serveurs il faudra utiliser la fonction `stream_set_timeout` pour ne pas bloquer le script en cas d'indisponibilité, cette non disponibilité ne signifiant pas en plus que le lien soit mort.

3.3 Requêtes et/ou actions à mener sur les champs de la base

Nous allons voir rapidement quelques exemples de modifications de la base.

3.3.a Compléter un champ dans toutes les références

EndNote sait remplacer des valeurs par d'autres valeurs dans un champ sur tous les enregistrements de la base ou insérer une valeur dans un champ vide mais est incapable de faire en une seule fois, l'ajout d'une valeur dans un champ vide et l'ajout de cette valeur précédée d'un retour à la ligne si le champ n'est pas vide. Voici un script pour ajouter automatiquement le nom de la base dans le champ `name_of_database` en première valeur si le champ n'est pas vide.

```
<?php
mysql_connect("localhost", "user", "passwd") or die ("Erreur de connexion");
mysql_select_db("rdb") or die ("Erreur de connexion à la base de donnée");
$requete='select id,name_of_database from refs order by id';
$result=mysql_query($requete) or die('Erreur SQL !'.$sql.'<br>'.mysql_error());
$name='nom_de_la_base.enl';
while ($data=mysql_fetch_array($result))
{
  if ($data[1]!="")
  {$ajout=$name.chr(13).$data[1];}
  else
  {$ajout=$name;}
  $rech='update refs set name_of_database="'.$ajout.'" where id='.$data[0];
  $resul=mysql_query($rech) or die('Erreur SQL !'.$sql.'<br>'.mysql_error());
}
echo 'Nom de la base ajouté';
mysql_close();
?>
```

A noter que si l'on veut ajouter une valeur à un champ sur la ligne suivante au cas où ce champ ne soit pas vide, il faudra toujours utiliser `chr(13)` dans le script pour faire un retour à la ligne `\r` et `\n` étant inopérants.

3.3.b Rechercher les références d'un ou plusieurs auteurs sans tirés à part sur le serveur local

Il suffit de donner le(s) nom(s) de(s) auteurs au script qui va chercher dans la base si les références dont les noms figurent dans le champ auteur possèdent un lien vers un tiré à part dans les champs URL et link_to_pdf vers le serveur local, en donnant une partie de son adresse. Le script fera un rapport donnant les id, correspondant aux record numbers des références concernées.

3.3.c Compléter le champ <Alternate Journal> à partir du champ <Journal> pour tous les journaux simultanément

Cette opération est légèrement plus complexe car elle fait intervenir deux bases. À noter qu'EndNote est capable de faire cette opération mais uniquement journal par journal. Dans notre cas, nous traitons tous les journaux avec une seule requête.

Nous possédons, sous la forme d'un fichier Excel les correspondances entre les titres complets et les titres abrégés. À partir de ce fichier, nous l'enregistrons au format csv, puis nous créons une base dans PHPMyAdmin avec des enregistrements de deux champs. Il suffit alors d'importer les données au format csv pour avoir une base nous permettant d'établir les correspondances.

En sélectionnant tous les champs secondary_title, correspondant aux titres des journaux, et en les comparant aux titres saisis dans la seconde base, on récupère le titre abrégé que l'on insère dans le champ alternate_title correspondant.

3.3.d Importer des références d'une base possédant les fichiers liés à la base

À partir de la version X, le répertoire Data peut avoir un sous répertoire PDF qui contient les fichiers de la base.

Dans le Menu Edit-Préférences vous choisissez l'option OpenURL/PDF. Si vous cochez l'option '*Copy new PDF Files to the default PDF folder and create a relative link*', le programme proposera par défaut de copier le fichier pdf et de faire un lien relatif lorsque vous ferez un Link to PDF dans une référence. Si vous ne cochez pas l'option, le programme proposera par défaut un lien absolu.

Dans le cas d'un lien absolu, celui-ci tient compte de la position absolue du fichier sur le disque à partir de la racine et dans le cas d'un lien relatif, le chemin a pour référence l'emplacement du fichier .enl de la base, le fichier pdf étant placé sous xxx.Data\PDF où xxx est le nom de la base. Si l'on examine les liens à partir de PHPMyAdmin, on constate que ces liens sont un peu particuliers car ils débutent tous par le mot internal.

Dans le cas de notre organisation, si on nous fournit une base avec des liens relatifs, il faut transformer tous ces liens en liens absolus, ce qui est long en utilisant EndNote.

Dans ce cas, il faut copier les références de la base que l'on vous fournit dans la base destinatrice et copier le répertoire PDF, toujours de la base que l'on vous fournit, au même niveau que les scripts en PHP. Vous créez ensuite un nouveau répertoire toujours au même niveau. Pour le traitement, vous sélectionnez les liens dont les 8 premiers caractères sont égaux à internal et vous récupérez le nom du fichier dans le lien. Il suffit ensuite de copier ce fichier dans le nouveau répertoire en le renommant (fonction copy de PHP) et de modifier le lien dans la base.

Les liens précédemment liés à la base deviennent des liens normaux. Vous déplacez ensuite les fichiers renommés à leur place habituelle et vous mettez la base à jour.

Ci-dessous, on constate que le script de transformation de la base est court et qu'il ne comporte aucune difficulté.

```
<?php
mysql_connect("localhost", "user", "password") or die ("Erreur de connexion");
mysql_select_db("rdb") or die ("Erreur de connexion à la base de donnée");
$requete='select id,link_to_pdf from refs where (left(link_to_pdf,4)!="http") and
(link_to_pdf!="")';
$result=mysql_query($requete) or die('Erreur SQL !'.$sql.'<br>'.mysql_error());
$ligne=mysql_numrows($result);
while ($data=mysql_fetch_array($result))
{
$num=$data['0'];
$fic=$data['1'];
if (substr($fic,0,8)=='internal')
{
$fic='PDF/'.substr($fic,15,strlen($fic)-15);
$nfic='puball/puball_'.$num.'.pdf';
if (!copy($fic,$nfic))
{echo $num.' '.$fic.'<br>';}
else
{
$lien='http://www.monsite.fr/bibliographie/tap
/prefixe_generique_'.$num.'.pdf';
$rech='update refs set link_to_pdf="'.$lien.'" where id='.$num;
$resul=mysql_query($rech) or die('Erreur SQL !'.$sql.'<br>'.mysql_error());
}
}
}
mysql_close();
?>
```

4. Organisation

Ces quelques exemples illustrent très partiellement ce que l'on peut réaliser facilement comme traitement.

Concrètement, un serveur web (EasyPHP) est implanté sur l'ordinateur servant à la gestion de la documentation sur lequel un petit site web a été mis en place offrant les options les plus couramment utilisées (**figure 4**). Lorsque la documentaliste veut réaliser une opération, elle recopie le répertoire rdb vers le répertoire de données d'EasyPHP, elle lance l'application

EasyPHP et donc le serveur web, elle donne l'adresse du site web local et elle exécute la ou les opérations voulues. Elle arrête le serveur en fermant EasyPHP.

Si la base a été modifiée, elle recopie le répertoire rdb vers le .Data de la base d'origine en tenant compte de ce qui a été dit au paragraphe 2 pour refaire le fichier .enl.

Hormis ces quelques options « standard », des opérations moins courantes ou plus complexes ont déjà été réalisées au cas par cas notamment lors de la fusion de différentes bases dont les informations saisies n'étaient pas homogènes.

Menu

Rappel : Il faut recopier le sous répertoire rdb situé sous .Data de la base EndNote à traiter dans le répertoire C:\Program Files\EasyPHP1-8\mysql\data
Attention !!! Il faut que la base ne soit pas ouverte dans EndNote lors de la copie.

- Edition de la liste des ouvrages (**base ouvrage.enl**).
- Publications d'un auteur par année et par type (**base publis_unite.enl**).
- Publications par année et par type (**base publis_unite.enl**).
- Ajout de la mention Fichier pdf dans le champ notes si une URL sur intranet existe (**base senso.enl**).
- Contrôle des liens URL et Link to pdf sur intranet d'une base (**toutes les bases**).
- Ajout des liens URL sur intranet d'une base (**toutes les bases**).
- Recherche des fichiers sans R.N. sur intranet pour une base (**toutes les bases**).
- Complément du champ Language (FRE ou ENG) pour une base (**toutes les bases**).
- Complément du champ Name of database pour une base (**toutes les bases**).
- Liste des journaux saisis dans une base (**toutes les bases**).
- Déplacement des url locales vers Link to pdf (**toutes les bases**).

Figure 4 : Menu du site web de la documentation

Conclusion et perspectives

EndNote permettant de gérer des données bibliographiques sous la forme d'une base, il est logique que ses concepteurs se soient appuyés sur un standard comme MySQL. Ce logiciel étant un produit fini, il est normal que certaines fonctions n'aient pas été implémentées par ses concepteurs, seules les opérations les plus fréquentes pour un utilisateur sont disponibles. Le fait de pouvoir utiliser et modifier les données de ce logiciel via des scripts PHP, qui sont en général simple à écrire, offre des possibilités quasi sans limites, hormis celle de ne jamais modifier la structure initiale de la base ce qui rendrait les données inutilisables par EndNote.

