



XML DBだからできること:HTMLをDBにそのまま保存

鵜飼寛太(サイバーテック)
2008-03-19 18:30:00

xhtmlで記述されたhtmlファイルを用意

最初に登録用htmlファイル(test.html)を用意する

```
<html>
<head>
  <meta http-equiv="Content-Type" content="text/html; charset=UTF-8" />
  <title>テストページ</title>
</head>
<body>
  <div>
    <h1>テストページ</h1>
    <div>htmlをそのままストアする</div>
  </div>
</body>
</html>
```

以下の点に注意すること。

- 文字コードのコンバートを省くためUTF-8で記述
- meta情報のcharsetは最初の日本語より前にないとPHPのloadHTMLFile関数で読み込んだ際に文字化けする。
- 通常のhtmlで記述されている場合、PHPのxmlパーサーの仕様でXML文書化されるので予期しない結果になることがある。

neoCore用関数を用意

次に以下のようなPHPスクリプトを用意する。

- PHPスクリプト([neoCore.functions.php](#))
- cURL API
 - 引数
 - httpリクエストメソッド: GETまたはPOST
 - neoCore接続先: neoadminまたはneoquery
 - 送信データ(省略化): httpリクエスト時の送信データ。通常はhoge=hoga&foo=barの形式
 - セッションID(省略可): getsession関数で取得したneoCoreセッションID
 - 戻り値
 - httpレスポンス: xml文書で返却されるので必要に応じて後処理を行う
- getsession(セッションの取得)
 - neoCoreコマンド: GETSESSION
 - パラメータ
 - ユーザ名

- パスワード(md5でハッシュされていること)
- 戻り値
 - セッションID:neoCoreからは<?xml version="1.0" encoding="UTF-8" ?><sid>Session ID</sid>で返却されるので値だけ取り出してく。
- endsession(セッションの破棄)
 - neoCoreコマンド: ENDESSION
 - 戻り値
 - 処理結果:neoCoreからは<?xml version="1.0" encoding="UTF-8" ?><success/>で返却されるので、必要に応じて後処理を行う。
- transactionstart(トランザクションの開始)
 - neoCoreコマンド: TRANSACTIONSTART
 - 引数: セッションID
 - 戻り値
 - 処理結果:neoCoreからは<?xml version="1.0" encoding="UTF-8" ?><success/>で返却されるので、必要に応じて後処理を行う
- transactioncommit(トランザクションのコミット)
 - neoCoreコマンド: TRANSACTIONCOMMIT
 - 引数
 - セッションID
 - 戻り値
 - 処理結果:neoCoreからは<?xml version="1.0" encoding="UTF-8" ?><success/>で返却されるので、必要に応じて後処理を行う
- transactionrollback(トランザクションのロールバック)
 - neoCoreコマンド: TRANSACTIONROLLBACK
 - 引数: セッションID
 - 戻り値
 - 処理結果:neoCoreからは<?xml version="1.0" encoding="UTF-8" ?><success/>で返却されるので、必要に応じて後処理を行う
- store(xml文書のストア)
 - neoCoreコマンド: STORE
 - 引数
 - セッションID
 - 格納するxml文書
 - プリフィックスファイル
 - スキーマファイル
 - 戻り値

- 処理結果:neoGoreからは

```
<?xml version="1.0" encoding="UTF-8" ?>
<Store-Results>
  <Documents-Processed> 1 </Documents-Processed>
  <Last-Doc-ID> 1199 </Last-Doc-ID>
</Store-Results>
```

で返却されるので、必要に応じて後処理を行う。



XML DBだからできること:HTMLをDBにそのまま保存

鵜飼寛太(サイバーテック)
2008-03-19 18:30:00

ストア用のPHPスクリプトを用意

対象のファイルが整形されたXML文書(xhtmlで記述されたファイル)の場合、次のスクリプトでneoCoreへストアできる。

```
<?php
/**
 *
 */
// neoCore関数のインクルード
include('neoCore.functions.php');
// neoCoreセッション開始
$cid = getsession("userid", md5("password"));
// xml文書の保存
$result = store($cid, file_get_contents("test.html"));
// neoCoreセッションの終了
endsession($cid);
//
# header('Content-type: text/xml');
// 結果の表示
# echo $result;
?>
```

※useridとpasswordは環境に合わせて適宜書き換えのこと

filegetcontentsの引数でストアするhtmlファイルを指定しているので、この引数を変更すれば他のhtmlファイルをストアできる。

実際はxhtmlで書かれていることは少ない(meta、br、input等が閉じられていない事が多い)ので念のためxhtml化し、加えてトランザクション管理を実装すると以下ようになる。

```
<?php
/**
 *
 */
// neoCore関数のインクルード
include('neoCore.functions.php');
// DOMDocumentのインスタンスを生成
$doc = new DOMDocument();
// 余計な空白を削除
$doc->preserveWhiteSpace = false;
// htmlファイルの読み込み
$doc->loadHTMLFile("test.html");
// XML文書化
$xml = $doc->saveXML();
// neoCoreセッション開始
$cid = getsession("userid", md5("password"));
// トランザクションの開始
transaction_start($cid);
// xml文書の保存
$result = store($cid, $xml);
// トランザクションの終了
transaction_commit($cid);
// neoCoreセッションの終了
endsession($cid);
//
header('Content-type: text/xml');
// 結果の表示
echo $result;
?>
```

[loadHTMLFileメソッド](#)で保存対象のhtmlをDOMDocumentにする。

これは[loadHTMLFileメソッド](#)(および[loadHtmlメソッド](#))の仕様で整形されたhtmlでなくてもパースできる事を利用している。

この際、妥当でない箇所はPHPのxmlパーサーの解釈によって補てんされるので意図しない結果になる可能性がある。



XML DBだからできること:HTMLをDBにそのまま保存

鵜飼寛太(サイバーテック)
2008-03-19 18:30:00

neoCoreの保存形式について

neoCoreへのストアが成功すると対象のXML文書のほかにneoCoreが付与するメタ情報が生成される。

```
<ND>
<MetaData>
  <ModifyTime>1200816236</ModifyTime>
  <TimeStamp>1200816236</TimeStamp>
  <SourceFile></SourceFile>
  <DocID>1199</DocID>
  <CopyNumber>1</CopyNumber>
</MetaData>
XML文書
</ND>
```

- neoCoreのメタ情報

- ModifyTime: 更新日時のタイムスタンプ
- TimeStamp: 生成日時のタイムスタンプ
- SourceFile: XML文書をファイル指定でストアした際のローカルファイル名
- DocID: neoCoreでドキュメントに付与されるユニークなID
- CopyNumber: DocID毎の版数

ストアはNDノードの子ノードとしてインサートされるので参照する際はストアしたXML文書のルートノードからのパスではなく、NDノードからのパスを書く必要がある。

この事を踏まえて、今回はストアしたXML文書を取りだすプログラムを作成する。



XML DBだからできること:HTMLをDBにそのまま保存

鵜飼寛太(サイバーテック)
2008-03-19 18:30:00

neoCore用関数をまとめたneoCore.functions.phpです。

```
<?php
/**
 *
 *
 * @package
 * @subpackage
 * @author Kanta Ukai <ukai@cybertech.co.jp>
 * @version
 */
/**
 * cURL
 * @param String HTTP Method GET|POST
 * @param String neoCore Command neoadmin|neoquery
 * @param String Data
 * @param String neoCore Session ID
 * @return String result
 */
function cURL ($method, $query, $data = null, $sid = null) {
    // cURLの初期化
    $ch = curl_init();
    //
    curl_setopt($ch, CURLOPT_HEADER, FALSE);
    curl_setopt($ch, CURLOPT_RETURNTRANSFER, TRUE);
    // セッションIDが設定されている場合、HTTPヘッダに設定する
    if ($sid) {
        //
        curl_setopt($ch, CURLOPT_HTTPHEADER, array('sid: ' . $sid));
        //
    }
    //
    $url = "http://localhost:7700/" . $query;
    // メソッド判定
    if ($method == 'GET') {
        // GETの場合、URLと連結
        $url .= '?' . $data;
        //
    } else {
        // POSTの場合、ポストフィールドに設定
        curl_setopt($ch, CURLOPT_POST, TRUE);
        curl_setopt($ch, CURLOPT_POSTFIELDS, $data);
        //
    }
    // 接続先の設定
    curl_setopt($ch, CURLOPT_URL, $url);
    // 実接続
    $result = curl_exec($ch);
    // cURLの終了
    curl_close($ch);
    //
    return $result;
}
/**
 * getsession
 * @param String neoCore User Name
 * @param String neoCore Password
 * @return String neoCore Session ID
 */
function getsession ($user = '', $password = '') {
    //
    $result = cURL('POST', "neoadmin", "cmd=GETSESSION&user={$user}&passwd={$password}");
    // <sid>SESSION ID</sid>で返却されるのでDOM化して取り出す
    $doc = DOMDocument::loadXML($result);
    $sid = $doc->documentElement->nodeValue;
    //
    return $sid;
}
/**
 * endsession
 * @param String neoCore Session ID
 * @return String Result XML
 */
function endsession ($sid) {
    //
```

```

    $result = cURL('GET', "neoadmin", "cmd=ENDSESSION", $sid);
    //
    return $result;
    //
}
/**
 * transaction_start
 * @param String neoCore Session ID
 * @return String Result XML
 */
function transaction_start ($sid) {
    //
    $result = cURL('GET', "neoquery", "cmd=TRANSACTION_START", $sid);
    //
    return $result;
    //
}
/**
 * transaction_commit
 * @param String neoCore Session ID
 * @return String Result XML
 */
function transaction_commit ($sid) {
    //
    $result = cURL('GET', "neoquery", "cmd=TRANSACTION_COMMIT", $sid);
    //
    return $result;
    //
}
/**
 * transaction_rollback
 * @param String neoCore Session ID
 * @return String Result XML
 */
function transaction_rollback ($sid) {
    //
    $result = cURL('GET', "neoquery", "cmd=TRANSACTION_ROLLBACK", $sid);
    //
    return $result;
    //
}
/**
 *
 */
function store ($sid, $xml, $prefix = "", $schema = "") {
    //
    if ($schema) {
        //
        $data['schemafilename'] = $schema;
        //
    }
    //
    if ($prefix) {
        //
        $data['prefixfile'] = $prefix;
        //
    }
    //
    $data['xmlsourcefile'] = $xml;
    //
    $result = cURL('POST', "neoquery?cmd=STORE", $data, $sid);
    //
    return $result;
    //
}
/**
 *
 */
function query ($sid, $input) {
    //
    $result = cURL('GET', "neoquery", "cmd=QUERY&input={$input}", $sid);
    //
    return $result;
    //
}

```