

「洋書の改版チェック」
&
「多言語版図書チェック」
を
OCLCのXISBNで行おう

平成24年6月26日

東京大学柏図書館 前田朗

コンテンツ

- 概論

- OCLCのxISBNサービス
- WebサービスとXML
- バッチ処理のとは
- バッチ処理の安全性

- サンプルスクリプトの利用

- サンプルスクリプトのソースコード
- サンプルコードの起動
- サンプルスクリプトコンソール
- サンプルスクリプト結果のみかた

概論

OCLCのXISBNサービス

- ◎ ISBNをキーに関連する版のISBNをまとめて得ることができるWebサービスです。
- ◎ ISBNを含む検索式をURLで与えてあげます
 - <http://xisbn.worldcat.org/webservices/xid/isbn/0465039138?method=getEditions&format=xml&fl=form,year,lang,ed>
- ◎ 検索結果は、XMLで得ることができます。

```
<rsp stat="ok">
  <isbn form="BC" year="2002" lang="eng" ed="[Nachdr.]">0465039138</isbn>
  <isbn form="BA" year="2001" lang="spa">8430604286</isbn>
  <isbn form="BA" year="2002" lang="chi" ed="Chu ban.">9867892453</isbn>
  <isbn form="BA" year="2004" lang="chi" ed="Di 1 ban.">7508600525</isbn>
  <isbn form="BA" year="2001" lang="jpn" ed="Shohan.">4881359932</isbn>
  <isbn form="BA" year="2001" lang="ger">3827004047</isbn>
  <isbn form="BA" year="1999" lang="eng" ed="3. print.">046503912X</isbn>
</rsp>
```

<http://current.ndl.go.jp/node/5440> も参照のこと

WEBサービスとXTML

◎ Webサービスとは

- Web上でコンピュータ向けに機能を提供するものです。
- Webサービスを複数組み合わせることもでき、これはマッシュアップといいます

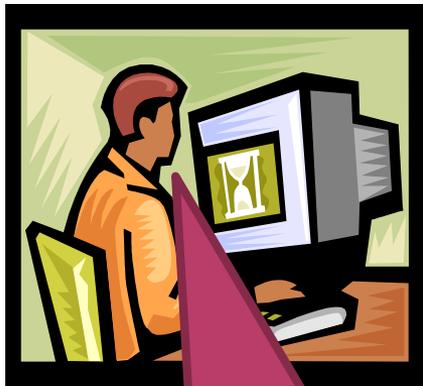
◎ XMLとは

- 標準的なデータ形式です。たとえば、書誌のように構造をもったデータをExcelの表で表現することは難しいですが、XMLであれば可能です。
- こういった標準がないと、プログラマは苦勞しあます。前田もかつて、個々に書式が違う特殊目録をデータベースに合わせ書式変換するプログラムをよく書いていましたが、なかなか大変でした。

バッチ処理とは

- ◎ バッチ（一括）処理は、簡単にいうと時間のかかる処理をコンピュータに仕掛けっぱなしにしておくものです。

リアルタイム処理



図書所蔵を1件ずつ
人力で検索

バッチ処理



図書所蔵を
100件まとめて
検索指示



バッチ処理の安全性

- ◎ バッチ処理は、コンピューターに処理をまとめて依頼します。実行中のトラブルにどうやって対処するのでしょうか。
 - 監視
 - コンソールに処理状況を垂れ流します。
 - オペレータがそれを監視します。
 - エラー処理
 - プログラムの中にエラー発生時の処理をきっちり埋め込んでおきます（リアルタイムのプログラムでもそうあるべきですが...）
 - ロット分割
 - エラーによる被害が広がらないように、処理データをあらかじめ、適切な分量（ロット）にわけます。
 - ロットの個数だけ、バッチ処理を実行します

サンプル スクリプト

このスクリプトの利用は自己責任でお願いします

サンプルスクリプトの ソースコード

```
#!/usr/bin/perl

use LWP::UserAgent;
use URI::Escape;
use XML::Simple;
use HTTP::Status;
binmode(STDOUT, ":utf8");

my $top_url = 'http://xisbn.worldcat.org/webservices/xid/isbn/';
my $tail_url = '?method=getEditions&format=xml&fl=form,year,lang,ed';
use constant TIMEOUT_DEF => 2;
use constant TIMEOUT_MAX => 10;
use constant LANG => chi; # eng => 英語, ger => ドイツ語, chi => 中文

my $i = 1;
while (my $isbn = <>) {
    chomp $isbn;
    my $url = $top_url . $isbn . $tail_url;

    print STDERR $i, "Yt", $url, "Yn";

    my ($content, $error) = _access_web($url, TIMEOUT_MAX);
    if ($error) {
        die "http error $error Yn";
    }
    print $i, "Yt", $isbn, "Yt";
    my $data = XMLin($content);

    if (ref $data->{isbn} eq 'ARRAY') {
        my @lang = grep { $_->{lang} eq LANG } @{$data->{isbn}};
        my @edit = sort { $b->{year} <=> $a->{year} } @lang;
        my $last = shift @edit;
        print $last->{ed}, "Yt",
              $last->{year}, "Yt",
              $last->{content}, "Yt",
              $last->{lang};
    }
    print "Yn";

    sleep(TIMEOUT_DEF);
    $i++;
}
```

```
sub disp_fields() {
    my $field = shift;
    my $item = shift;
    print $item->{"$field"}, "Yt"
        unless ref $item->{"$field"} eq "HASH";
}

# Webサーバへのアクセス
sub _access_web {
    my $url = shift;
    my $timeout = shift;
    my $usa = new LWP::UserAgent;
    my $request = new HTTP::Request('GET', $url);
    $usa->timeout([$timeout]);

    my $response = $usa->request($request);
    if ($response->code != RC_OK) {
        return (undef, $response->code);
    }
    else {
        return ($response->content, undef);
    }
}

__END__

Copyright Maeda Akira (maeda@lib.u-tokyo.ac.jp)
```

Perlで書いたコードです。Windows Linuxそれぞれで動きます

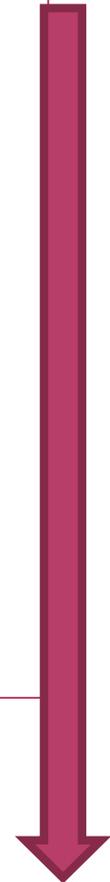
サンプルスクリプトの起動

- ◉ 前ページのサンプルコードを、`xisbn.pl`というファイルに保存します。
 - デフォルトは「英語の最新版のISBNを返す」設定ですが、`LALG`パラメータを直すことで、ドイツ語の最新版、中文の最新版等の変更ができます。
- ◉ ISBNのリスト（ISBNを改行で区切ったもの）を`isbnlist.txt`というファイルに保存します。
- ◉ Linuxのシェルもしくは、Windowsのコマンドプロンプトで上記の2つのファイルを置いたディレクトリに移動します。
- ◉ 次のコマンドを実行します。
 - `Perl xisbn.pl isbnlit.txt > isbnreslut.txt`

サンプルスクリプトコンソール

- 1 <http://xisbn.worldcat.org/webservices/xid/isbn/0674443020?method=getEditions&format=xml&fl=form,year,lang,ed>
- 2 <http://xisbn.worldcat.org/webservices/xid/isbn/0198293569?method=getEditions&format=xml&fl=form,year,lang,ed>
- 3 <http://xisbn.worldcat.org/webservices/xid/isbn/0375405445?method=getEditions&format=xml&fl=form,year,lang,ed>
- 4 <http://xisbn.worldcat.org/webservices/xid/isbn/006019314X?method=getEditions&format=xml&fl=form,year,lang,ed>
- 5 <http://xisbn.worldcat.org/webservices/xid/isbn/0195118049?method=getEditions&format=xml&fl=form,year,lang,ed>
- 6 <http://xisbn.worldcat.org/webservices/xid/isbn/0029328802?method=getEditions&format=xml&fl=form,year,lang,ed>
- 7 <http://xisbn.worldcat.org/webservices/xid/isbn/0710211716?method=getEditions&format=xml&fl=form,year,lang,ed>

現在の処理件数と実行中の検索式を
モニタリングできます



サンプルスクリプト結果のみかた

- ◎ コンソールにテキストが流れなくなったら、処理が終了です。
- ◎ エディタで、isbnresult.txtを開きます。
- ◎ 指定した言語の最新版があれば、次のように表示されます。

67

68

69 2. Aufl. 1995 3527293426 ger

70

71