



Technical Note 05-18

cURL を使って web フォームに自動入力

By Thomas Maul, General Manager, 4D Germany.
Technical Note 05-18

(原題: cURL - HTTP Client, Get and Post, FTP and much more.)

概要

cURL は、Windows ではインストール不要、MacOS X も最初から利用できるオープンソースのツールです。HTTP(s)を使用して GET、POST を実行することができ、クッキー、リファラー、ユーザーエージェント、認証などの機能を利用することができます。FTP その他のプロトコルを使用することができ、フォームの自動入力や送受信の機能も備えています。今回は 4D 2004 のコマンド LAUNCH EXTERNAL PROCESS を使用して cURL を制御する方法をご紹介します。

cURL とは？

<http://curl.haxx.se>

cURL のweb サイトでは次のように説明されています。「cURLとはURL シンタックスでファイルを転送するためのコマンドラインツールで、FTP、FTPS、HTTP、GOPHER、TELNET、DICT、FILE およびLDAP をサポートしている。HTTPS 認証、HTTP POST、HTTP PUT、FTP アップロード、HTP フォームによるアップロード、プロキシ、クッキー、パスワード認証 (Basic, Digest, NTLM, Negotiate, kerberos...)、ファイル転送のレジューム機能、プロキシタンネリング、その他、無数の便利な機能を有している。」cURL はオープンソースで、MacOS X では、出荷時にシステムに含まれています。

ターミナルを起動して、以下のように入力してみればcURLを試してみることができます。

```
curl www.4D.com
```

Windows の場合は、別途、ファイルを用意しなければなりませんが、4D データベースストラクチャと同じ階層にコピーするだけでよく、インストール作業は簡単です。新しい4D Extrasフォルダを使用すれば、すべてのクライアントマシンにcURLを転送することができます。ライセンス規約の「Installation on Windows」セクションをみると、このような使用が許可されています。

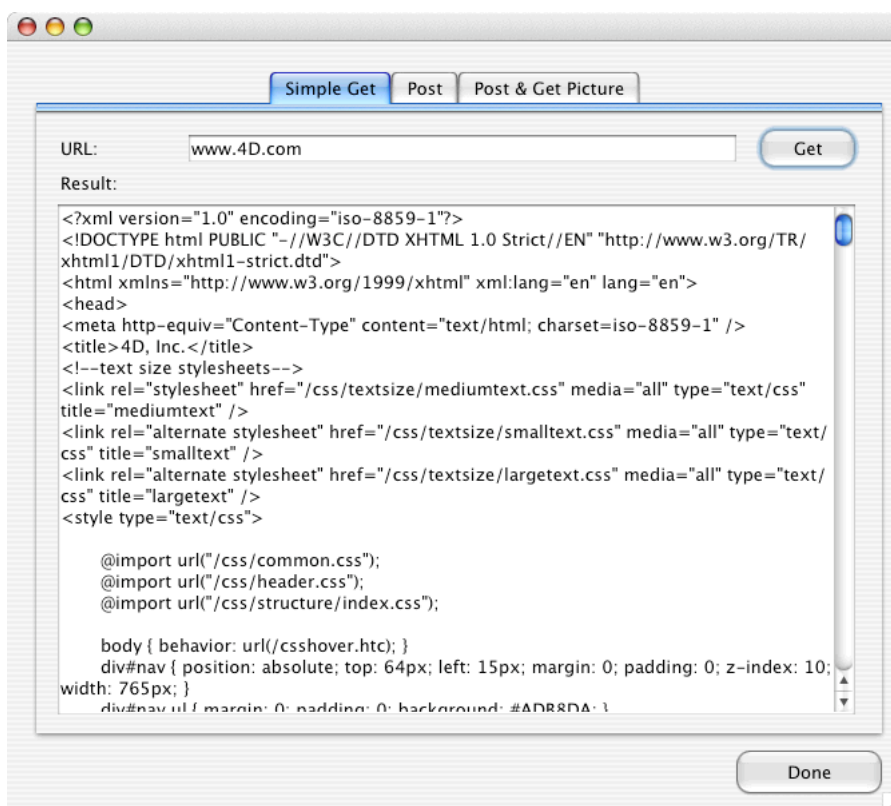
加えてcURL はライブラリとしても入手可能で、これを利用した4D プラグインをコンパイルすることも可能です。ここではコマンドLAUNCH EXTERNAL PROCESSを使用し、プラグインを開発せずにcURLを利用する方法に焦点を当てています。この方法の利点は、バージョンアップに柔軟に対応できることです。これに対してプラグインの場合は、再コン

ファイルが必要です。LAUNCH EXTERNAL PROCESS は4D のバージョン変更、OS のバージョン変更、その他コンポーネントの変更による修正を最小限にとどめることができます。

サンプルデータベース

サンプルデータベースを起動してみれば、cURLの概要がつかめると思います。用意するのは4D 2004.2 およびインターネット接続だけです。ストラクチャはMacOS XおよびWindows で使用することができ、Windows 用のファイルも含まれています。Quicktime だけは別途インストールしておいてください。

最初のタブは簡単なGET を実行するデモです。適当なURL を入力し、GET ボタンをクリックすると、ページがダウンロードされ、テキストのみの内容が表示されます。



注記:

デモを簡単にとどめるために、あえてプロキシには対応させませんでした。cURL自体は、プロキシに対応しており、たとえばこのデモで-x IP:port www.4D.com と入力すればプロキシを通して接続することができます。

ボタンのオブジェクトメソッドは以下のとおりです。

`vResult:=RunCurl (vURL)`

RunCurl はcurl をコールするためのプロジェクトメソッドで、MacOS もしくはWindows による方法の違いを処理し、結果をテキストで返します。ちなみにRunCurlBlob というプ

プロジェクトメソッドも用意されており、こちらは32kb 以上の結果、たとえば画像やPDF 文書などをBLOB 形式で返すために使用します。

次はもう少し高度な処理をしてみましょう。タブコントロールをページ2 に移すと、今度はHTTP POST のデモが表示されます。

これはヨーロッパで実際の請求業務に使用できる実用的なサンプルです。EU 内の企業は、それぞれVAT（付加価値税）番号というものを割り当てられており、互いに有効なVAT 番号を持つもの同士では出荷にかかるVAT が免除になる制度がとられています。したがって、発送者は相手のVAT が有効かどうかを確認する必要があるのですが、通常この業務は専用のweb ページを開いてブラウザに入力することで行ないます。

このデモではその処理を自動化、つまり必要事項をWebページに自動的に入力し、結果を解析しています。

最初の例よりは処理が複雑ですが、それでも受信HTML を解析するのではなく、知りたい情報、すなわちnumber is validとnumber is not validをパラメータとして使用している点に注目することができます。

実際のweb ページをブラウザで開いてみると、仕組みを理解するための助けになります。

http://www.europa.eu.int/comm/taxation_customs/vies/en/vieshome.htm

ブラウザは、便利な機能が用意されているのでFirefox/Mozillaがおすすめです。

EUROPA site - Validation

http://europa.eu.int/comm/taxation_ci

はじめよう 最新ニュース

Important legal notice

English (en)

EUROPA > European Commission > Taxation and Customs Union > VIES

Contact | Search on EUROPA

What's new? | Key issues | Information | FAQ

VIES VAT number validation

You can verify the validity of a VAT number issued by any Member State by selecting that Member State from the drop-down menu provided, and entering the number to be validated.

Member State

VAT Number

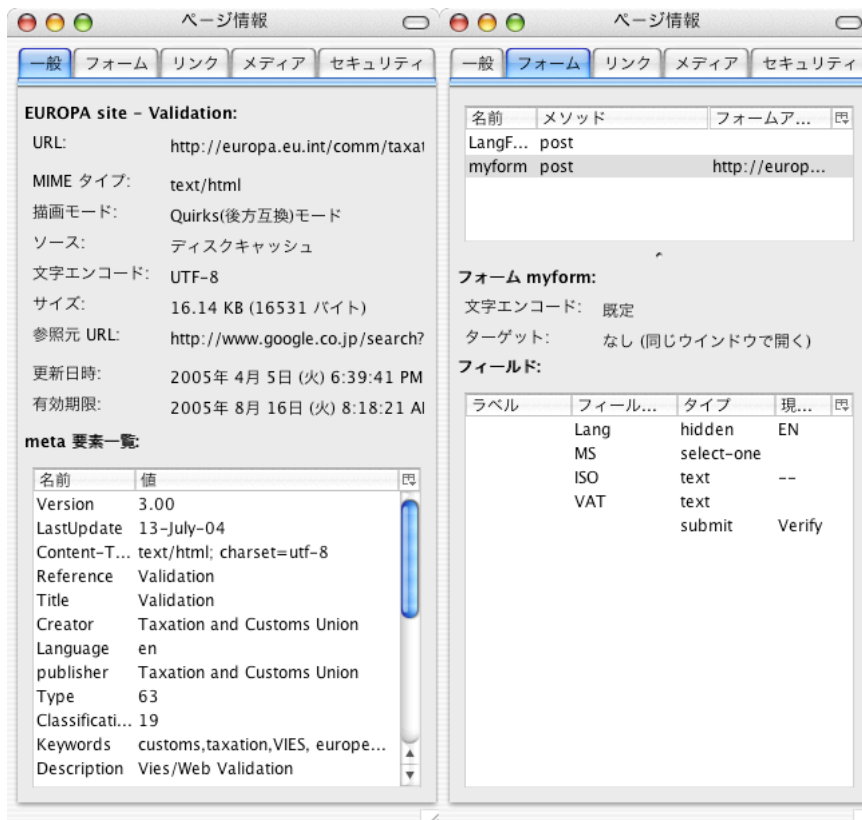
Current Date 04/08/2005 (dd/mm/yyyy)

Verify

FAQ | Help | Specific disclaimer for this service

完了

ブラウザでページを開くと、入力フィールドが3 つと、確定ボタンがひとつあることに気づきます。フォームに自動入力するには、フォーム名が必要です。ソースをみても構いませんが、Firefox にはもっと簡単な方法が用意されています。右クリックもしくはコンテキストメニューより「ページ情報をみる」を選択すると、フォームの情報が表示されます。



フォームタブをクリックすると、ページにはフォームがふたつあることが分かります。2番目のフォームを選択すると、3つの入力フィールドだけでなく、タイプhiddenのフィールドがあり、その内容はweb ページの言語情報であることが確認できます。

HTML POST には、各フィールドの値を&記号でつなげて記述するという規則があります。

Lang=EN&MS=DE&ISO=DE&VAT=128943976

Curl はこのシンタックスを自動的に作成してくれます。POST ステートメントであることを明示するためには、"-d" オプションを加えることになっています。ステートメント自体は引用符 (char(34)) で囲みます。

```
$post:"-d "+Char(34)+"Lang=EN&MS="+$iso+"&ISO="+$iso+"&VAT="+Substring(vVAT;3)+Char(34)
```

同じことを他のサイトでも試してみましょう。Google ではイメージ検索が可能ですが、実行には、検索パラメータ、クッキーの処理、有効なリファラーおよびユーザーエージェントが必要で、ない場合はサービスが拒否されてしまいます。このような制限を持つページは少なくないので、Google はその仕組みを理解する良い材料だといえるでしょう。

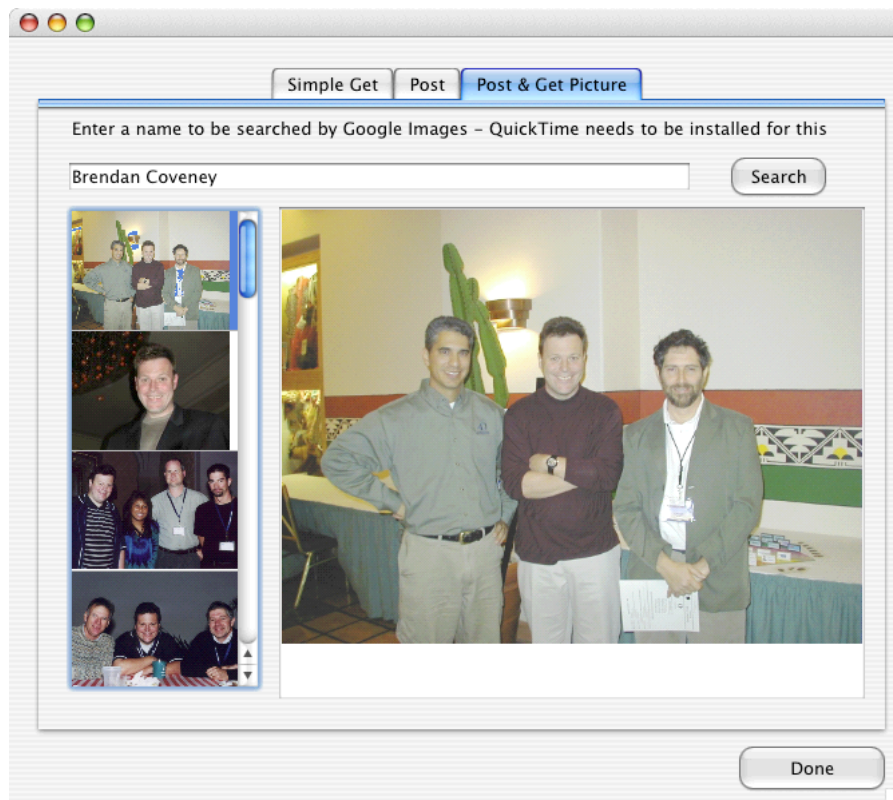
注記:

Google のライセンスは自動アクションを認めていません。このコードは高度な自動検索を実行する方法を示すためのもので、特にGoogle を対象にする意図はありません。

検索フィールドにキーワードを入力し、Search ボタンをクリックすると、最初にヒットした画像が20 件表示されます。おおまかに分けると次のような処理が行なわれています。

mage.google.com から有効なクッキーを取得する

フォームに値を入力、送信、結果を受信する
結果を分析し、画像のURL を取得する
URL から元の画像を取得する



クッキーを取得する方法

Curl は、次のオプションを使用することによってクッキーを送信することができます。

```
"-b mycookie=Hello"
```

"-i" オプションを使用して、ヘッダ全文を受信すれば、クッキーを受信することもできます。ただし、この場合、ヘッダの解析は各自で行なわなければなりません。それよりもCurl の素晴らしいところは、受信したクッキーをすべてテキスト形式で保存し、将来のリクエストに活用できるという点です。

```
$url="-b cookies.txt -c cookies.txt http://images.google.com/"  
vResult:=RunCurl ($url)
```

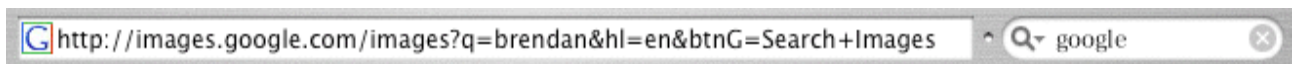
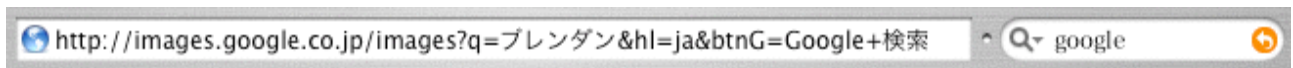
オプション"b"は、ファイルcookies.txtからすべてのクッキーを自動的に送信します。オプション"-c"は、ヘッダを解析し、クッキーを取り出して自動的に同ファイルにそれを保存します。したがって、クッキーの処理はすべてプログラムに任せてしまうことができます。

フォームに入力して結果を取得

基本的にはVAT のデモと同じですが、さらに以下の要素を考慮する必要があります。

エンコード

例では、キーワードの区切りに半角スペースを用いました。半角スペースは直接送信することができないので、エンコードする必要があります。同じことは、7 ビット文字の一部と、それ以上のすべての文字、たとえばô、é などについてもいえます。エンコードには、文字のISO-8859-1 コードと&が使用されます。実際にブラウザでGoogle 画像検索を実行してみて、URL をみてみましょう。目標は、このURLエンコードを再現することです。



リファラーとユーザーエージェント

cURLは、オプション”A”でユーザーエージェントを作成することができます。ここでは、Windows95 でMicrosoft Internet Explorer 4.0 を使用しているようにサーバに認識させました。オプション”b”はクッキーを、”e”はリファラーを送信します。

注記:

cURL のオプション引数は、小文字と大文字では意味が異なります。

変数vResult には、画像のURL を含むHTML 文書が格納されます。今回の例では、完璧なパーサーを用意する必要はなく、目印となる目印すなわち<a href=/imgres?imgurl=だけに注目して画像のURLを取得するだけで充分です。

画像のURLが配列arPictURL に格納されたら、あとはダウンロードをするだけですが、ポイントはボタンのオブジェクトメソッドでそこまでを記述していない点です。20 枚の画像を20 基のサーバからダウンロードするには、かなりの時間を要するかもしれず、インタフェースが停止したかのような印象を与える恐れがあります。画像を一枚ダウンロードするたびに表示してゆけば、少なくとも何かが実行されているという印象をユーザに与えることができるので、そのためにSET TIMER をフォームメソッドで使用しています。

次のステップ

このサンプルはcURLが提供する多彩な機能のほんの一部を紹介したに過ぎません。cURLの詳細についてはマニュアルを参照してください。

<http://curl.mirror.internet.tp/docs/manual.html>

例題

<http://curl.mirror.internet.tp/docs/manpage.html>

<http://curl.mirror.internet.tp/docs/httpscripting.html>

ブラウザとサーバの間の通信については以下のサイトが参考になると思います。

<http://livehttpheaders.mozdev.org/>

cURLによる自動処理がうまくいかない場合、原因のほとんどはクッキー、リファラーなどの処理が適切でないことにあります。cURLとWebサーバの間のやり取りはみえないため、こうした原因を突き止めるのは大変かもしれません。パケットスニッファで通信をたどる方法もありますが、データ量が多いとかえって混乱するので注意が必要です。

Windows/MacOS 用: Ethereal

<http://www.ethereal.com/>

MacOS X 用: MacSniffer

http://personalpages.tds.net/~brian_hill/macsniffer.html

"HTTPLook"のような製品は、ブラウザ、cURL、4D Internet Commands などすべてのHTTPパケットのログを保存するので、こうした問題の特定には最適のツールだといえます。

Windows 用: <http://www.httpsniffer.com/> (デモ版あり)

Windowsマシンにインストールする

cURLは、普通のexeファイルで、DLLやレジストリエントリなどをしなくても実行することができます。4D Clientマシンへ配布するには4D Extrasフォルダの使用がおすすめです。そうすれば、次のコードでパスを取得することができます。

```
$path:=Get 4D folder(Extras Folder )
```

```
$path:=$path+"Curl.exe"
```

なお、サンプルに同梱してあるのは、version 7.13.2 Non SSLです。最新バージョンは以下のサイトで確認することができます。

<http://curl.mirror.internet.tp/download.html>

SSLバージョンを利用するためには、OpenSSLをインストールする必要があります。OpenSSLはオープンソースで、cURLと同じページからダウンロードできますが、再配布するにはライセンス事項に留意しなくてはなりません。(ライセンス、マニュアル、Aboutダイアログにコメントを追加すること)

LDAPなどの機能を使用するためには、別途、ライブラリ等を用意します。