

Web Automation

Using Perl to automate web surfing

Why Automate Surfing?

The automatch program

```
#!/usr/bin/perl -w

# The 'automatch' program - check a collection of sequences against
# the 'mersearchmulti.html' web page.

use strict;

use constant URL => "http://pblinux.itcarlow.ie/mersearchmulti.html";

use WWW::Mechanize;

my $browser = WWW::Mechanize->new;

while ( my $seq = <> )
{
    chomp( $seq );

    print "Now processing: '$seq'.\n";
```

The automatch program, cont.

```
$browser->get( URL );
$browser->form( 1 );
$browser->field( "shortsequence", $seq );
$browser->submit;

if ( $browser->success )
{
    my $content = $browser->content;

    while ( $content =~
        m[<tr align="CENTER" /><td>(\w+?)</td><td>yes</td>]g )
    {
        print "\tAccession code: $1 matched '$seq'.\n";
    }
}
else
{
    print "Something went wrong: HTTP status code: ",
        $browser->status, "\n";
}
}
```

Running the automatch program

```
$ chmod +x automatch
```

```
$ ./automatch sequences.txt
```

Results from automatch ...

```
Now processing: 'attccgattagggcgta'.
Now processing: 'aattc'.
  Accession code: AF213017 matched 'aattc'.
  Accession code: J01730 matched 'aattc'.
  Accession code: M24940 matched 'aattc'.
Now processing: 'aatgggc'.
Now processing: 'aaattt'.
  Accession code: AF213017 matched 'aaattt'.
  Accession code: J01730 matched 'aaattt'.
  Accession code: M24940 matched 'aaattt'.
Now processing: 'acgatccgcaagtagcaacc'.
  Accession code: M15049 matched 'acgatccgcaagtagcaacc'.
Now processing: 'gggcccaaa'.
Now processing: 'atcgatcg'.
Now processing: 'tcatgcacctgatgaacgtgcaaaaccacag'.
  Accession code: AF213017 matched 'tcatgcacctgatgaacgtgcaaaaccacag'.
  .
  .
Now processing: 'ccaaat'.
  Accession code: AF213017 matched 'ccaaat'.
  Accession code: J01730 matched 'ccaaat'.
  Accession code: M24940 matched 'ccaaat'.
```

Viewing the source of the mersearchmulti.html web page



```
<html>
<head>
  <title>Search the "dnas" Table for a Match</title>
</head>
<body>

Please enter a sequence to match against the database:<p>
<form ACTION="/cgi-bin/db_match_emblCGI">
  <p>
    <textarea name="shortsequence" rows="4" cols="60"></textarea>
  </p>
  <p>
    Include a border around the results: <input type="checkbox"
name="printborder" value="on" checked="checked" />
  </p>
  <p>
    <input type="reset" value="Clear">
    <input type="submit" value="Try it!">
  </p>
</form>
</body>
</html>
```

Maxim 16.1

Automate repetitive WWW interactions
whenever possible

Where To From Here