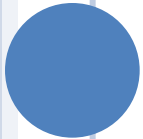# Homework 01

Announce: 20090325

Due: 20090401

Requirements

- Use Perl with CPAN modules to build a web proxy with record feature
- Use the logs your recorded to turn web applications to CIL application
  - With batch and addition features!
- Example
  - Dictionary/Wiki lookup
  - Search on multiple search engines
  - Album grabber
  - Auto register
  - etc.

Proxy

- HTTP::Proxy
  - /usr/ports/www/p5-HTTP-Proxy
  - http://search.cpan.org/dist/HTTP-Proxy/
- HTTP::Recorder
  - /usr/ports/www/p5-HTTP-Recoder
  - http://search.cpan.org/dist/HTTP-Recorder/

  - http://http-recorder/

Example Code

```perl
use HTTP::Proxy;
use HTTP::Recorder;

my $proxy = HTTP::Proxy->new(
                port => 3128,
                host => undef);


my $agent = new HTTP::Recorder;
$agent->file("log");

$proxy->agent( $agent );
$proxy->start();
```

4

Set Proxy

Get code!

```
$agent->get('http://www.google.com/dictionary');

$agent->form_name('f');
$agent->field('q', 'Serendipity');
$agent->field('langpair', 'en|zh-TW');
$agent->click();
```

Bot

- WWW::Mechanize
  - /usr/ports/www/p5-WWW-Mechanize
  - http://search.cpan.org/dist/WWW-Mechanize/

Example Code

```
use WWW::Mechanize;
my $agent = WWW::Mechanize->new();


#
# Paste and modify what you recorded here
#
# $agent-> …
# …
#
```

Other CPAN modules

- User Interface
  - devel/p5-Curses
    - devel/p5-Curses-UI
    - devel/p5-Curses-*
  - devel/p5-Dialog
- Parallelization
  - www/p5-ParallelUA
- Cookies
  - www/p5-libwww
    - my $cookie = HTTP::Cookies->new();
    - my $m = WWW::Mechanize->new(
                                    cookie_jar => $cookie);

FAQ

- "Parsing of undecoded UTF-8 will give garbage when decoding entities at /usr/local/lib/perl5/site_perl/5.8.9/mach/HTML/PullParser.pm line 81."
  - use utf8;
  - Set all your environment to UTF-8
- HTTP::Recorder doesn't provide enough information
  - http://search.cpan.org/dist/WWW-Mechanize/lib/WWW/Mechanize.pm
    - LINK METHODS
    - IMAGE METHODS
      - find_*()