WebVoyáge with a Wrapper

Michael Doran, Systems Librarian
doran@uta.edu

Ex Libris Users of North America (ELUNA) Meeting
Long Beach, CA - Session 48.2 - July 31, 2008

Once upon a time...

"Hi, my name is Lloyd"

"If you will just configure me right, I will turn into a handsome OPAC."

What is a handsome OPAC?

- Aesthetically handsome
- Functionally handsome
- An OPAC is “handsome” if it
  - is simple to use
  - is intuitive to use
  - makes it easy to find stuff
- “Only librarians like to search, everybody else likes to find.”
  – Roy Tennant

Simple searches

- Google
- Amazon
- eBay

WebVoyáge simple search

The secret handshakes

- last name, first name for author searches
- no initial articles for title searches
- Library of Congress subject headings
- Boolean operators
- what an index browse is
**WebVoyáge simple search (after)**

- keyword anywhere search
- words within quotes are treated as a phrase
- other words are automatically Boolean ANDed*
- relevancy ranked results*

Hmmm… that’s Google-like functionality

---

**WebVoyáge server-side back end**

![Diagram of WebVoyage server-side back end]

**WebVoyáge is a “black box”**

![Diagram of WebVoyage as a black box]

**They call it a wrapper**

![Diagram of WebVoyage as a wrapper]

**Basic wrapper script**

```perl
#!/usr/bin/perl
$data_stream = `./Pwebrecon-orig.cgi`; print $data_stream; exit;
```

---

**Code**

![Code snippet for WebVoyage]

---

Michael Doran, Systems Librarian
doran@uta.edu
Do your thing to that datastream

- aka “screen scraping”
  “A technique in which a computer program extracts data from the display output of another program. The key element that distinguishes screen scraping from regular parsing is that the output being scraped was intended for final display to a human user, rather than as input to another program, and is therefore usually neither documented nor structured for convenient parsing.” [from Wikipedia]

- text wrangling
  add text
  delete text
  rearrange text

Example – adding text

```
#!/usr/bin/perl
$data_stream = `./Pwebrecon-orig.cgi`
$data_code = qw(<link rel="stylesheet" type="text/css" href="/css/my.css">);
$data_stream =~ s#</HEAD>#$meta_code</HEAD>#$1;
print $data_stream;
exit;

<HTML>
<HEAD>
<meta http-equiv="Content-Type" Content="text/html; charset=UTF-8">
<TITLE>Library Catalog - University of Texas at Arlington</TITLE>
<link rel="stylesheet" type="text/css" href="/css/my.css">;
</HEAD>
</BODY onLoad="document.querybox.Search_Arg.focus()"
```

Example – removing text

```
#!/usr/bin/perl
$data_stream = `./Pwebrecon-orig.cgi`
$data_stream =~ s#<TD><STRONG>.*?</STRONG>University of Texas at Arlington Library</TD>##i;
print $data_stream;
exit;
```

Example – rearranging text

```
#!/usr/bin/perl
$data_stream = `./Pwebrecon-orig.cgi`
$data_stream =~ s#Search request: (.*)</TD>.*Results: (.*).
entries.#<br />$2 entries for <b>$1</b>#
print $data_stream;
exit;
```

Wrapper script redux

```
#!/usr/bin/perl
$data_stream = `./Pwebrecon-orig.cgi`
print $data_stream;
exit;
```

- Read and parse form input
  - QUERY_STRING (get method)
  - STDIN (post method)

---

Michael Doran, Systems Librarian
doran@uta.edu
Truncation adaptation

Incoming data

Example – truncation adaptation

Other input data munging

- fix Voyager 6.x GKEY/TKEY/SKEY keyword "multiple spaces" no hits bug (Support Web incident #131344)
- deal with "right single quotation mark" vs. "apostrophe" in search input issue
- allow for ISBNs with dashes*
- (combined output/input) data munging
Is a wrapper right for you?
- requires some programming expertise
- requires lots (and lots) of testing
  - test platform
    - ideally a Voyager test server
    - separate WebVoyage instance (a la preview server)
    - law of unintended consequences
- extra layer makes WebVoyage more brittle
  - more dependencies, e.g. with opac.ini
  - upgrades more complicated

Getting started
- wrappers are programming-language-neutral, however…
- Perl is good
  - designed for text processing
  - robust regular expressions
  - is already on your system
  - example wrappers available
  - it’s fine to think big…
  - … but start small

Resources
- Michael Doran, University of Texas at Arlington
  Presentation: “WebVoyage with a Wrapper”
  Source code: http://rocky.uta.edu/doran/wrapper/
- Ere Maijala, National Library of Finland
  European EndUser 2006 presentation: “Enhancement scripts for WebVoyage OPAC”
  (no longer on the web – Contact Ere Maijala)
  Source code: http://www.nationallibrary.fi/libraries/linnea/pwebrecon2.html

A small start
- copy original Pwebrecon.cgi
  cp -p Pwebrecon.cgi Pwebrecon-orig.cgi
- create Pwebrecon.cgi wrapper template
  #!/usr/bin/perl
  $data_stream = `./Pwebrecon-orig.cgi`;
  print $data_stream;
  exit;

  - add desired feature
  - test

Will wrappers work with Tomcat?
- Is it possible to apply a wrapper to the Voyager 7 Tomcat WebVoyage?
- Is it desirable to apply a wrapper to the Voyager 7 Tomcat WebVoyage?
  - V7 Tomcat XML files
  - V7 Cascading Style Sheets

Will wrappers work with Tomcat?
- Is it possible to apply a wrapper to the Voyager 7 Tomcat WebVoyage?
  - No “Pwebrecon.cgi” under Tomcat WebVoyage (in fact no CGI period)
  - No one single controlling executable to wrap