

Homework 6: Server-Side Scripting in Perl

1. Objectives

- Get experience with Perl
- Get experience with CGI programming
- Get experience with XML

2. Requirements

In this exercise, you are required to create a web site, which will provide to a visitor the ability to search the current TV listings. When a user selects a given channel, your script will retrieve its shows and other information (name, time, link, etc.) and display it.

A user will first connect to a web page that contains a form containing the following element:

- A button labeled “Collect Data from <http://tv.yahoo.com/listings>”

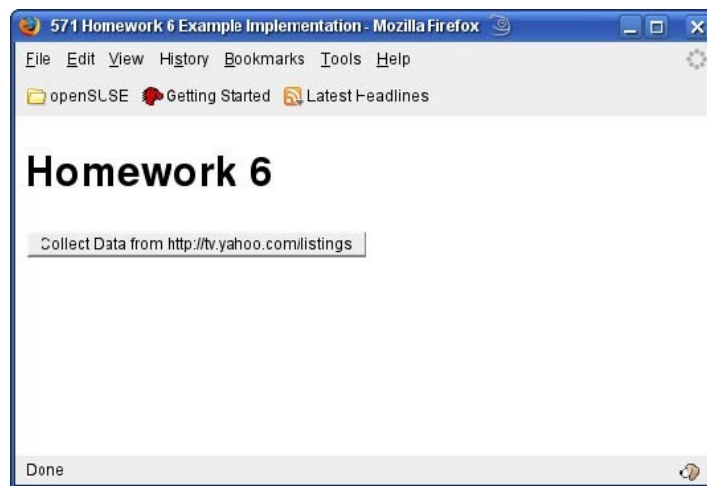


Figure 1

2.1 When "Collect Data from <http://tv.yahoo.com/listings>" is clicked (See Figure 1), your web page will invoke a server-side script in Perl that will do the following:

- Produce an XML file that contains the details of all current channels listed and all their shows.

2.2 Next a new page will be displayed that includes the following hyperlinks:

1. [Search...](#)
2. [Yahoo! TV listings\(XML\)...](#)



Figure 2

For each channel the Yahoo! TV listings website <http://tv.yahoo.com/listings> displays a list of shows

and and the associated time slots. In addition, it may or may not contain a link to a website where more information can be found about a particular show. Figure 3 shows an example of the Yahoo TV listings website which you would have to “scrape” to extract the information about the channels, shows, etc.

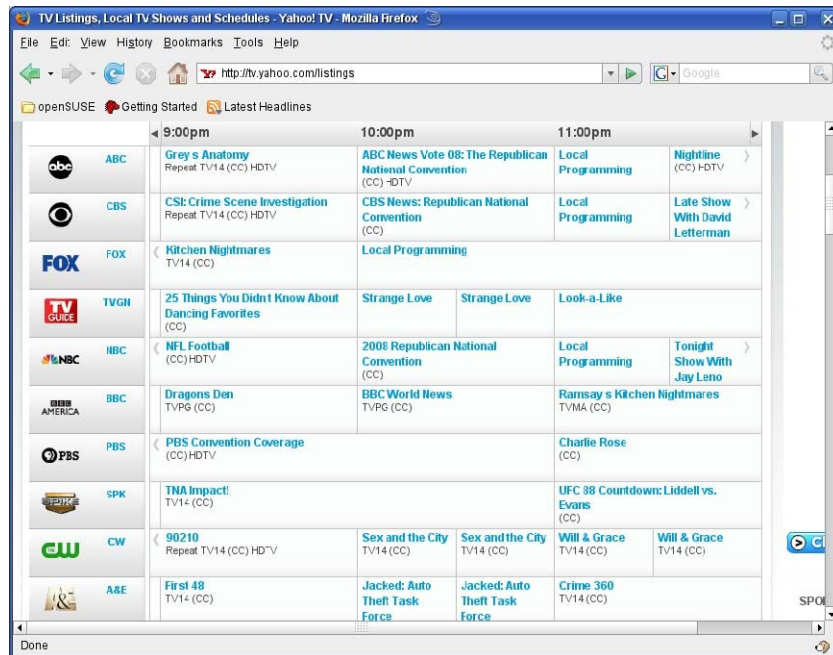


Figure 3

Your Perl script will extract the information and produce an XML file with the following format.

```
<?xml version="1.0" encoding="UTF-8"?>
<TVLISTINGSRESULTS>
  <CHANNELS>
    <CHANNEL>
      <NAME> ABC </NAME>
      <SHOW>
        <TIMESLOT> 3:00pm -4:00pm </TIMESLOT>
        <LINK>
          http://tv.yahoo.com/listings/general-hospital/show/86
        </LINK>
        <SHOWNAME> General Hospital </SHOWNAME>
      </SHOW>
      <SHOW>
        <TIMESLOT> 4:00pm -6:30pm </TIMESLOT>
        <LINK> N/A </LINK>
        <SHOWNAME> Local Programming </SHOWNAME>
      </SHOW>
    </CHANNEL>
    <CHANNEL>
      <NAME> CBS </NAME>
      <SHOW> <TIMESLOT> 3:00pm -4:00pm </TIMESLOT>
      <LINK>
        http://tv.yahoo.com/listings/guiding-light/show/30848
      </LINK>
      <SHOWNAME> Guiding Light </SHOWNAME>
```

```

        </SHOW>
        <SHOW>
            <TIMESLOT> 4:00pm -6:00pm </TIMESLOT>
            <LINK> N/A </LINK>
            <SHOWNAME> Local Programming </SHOWNAME>
        </SHOW>
    </CHANNEL>
    <CHANNEL>
        <NAME> FOX </NAME>
        <SHOW>
            <TIMESLOT> 3:00pm -6:00pm </TIMESLOT>
            <LINK> N/A </LINK>
            <SHOWNAME> Local Programming </SHOWNAME>
        </SHOW>
    </CHANNEL>
    <CHANNEL>
        <NAME> TVGN </NAME>
        <SHOW>
            <TIMESLOT> 3:00pm -4:00pm </TIMESLOT>
            <LINK> N/A </LINK>
            <SHOWNAME> InFANity </SHOWNAME>
        </SHOW>
        <SHOW>
            <TIMESLOT> 4:00pm -5:00pm </TIMESLOT>
            <LINK> N/A </LINK>
            <SHOWNAME> Hollywood 411 </SHOWNAME>
        </SHOW>
        <SHOW>
            <TIMESLOT> 5:00pm -5:30pm </TIMESLOT>
            <LINK> N/A </LINK>
            <SHOWNAME> Hollywood 411 </SHOWNAME>
        </SHOW>
        <SHOW>
            <TIMESLOT> 5:30pm -6:30pm </TIMESLOT>
            <LINK>
                <a href="http://tv.yahoo.com/listings/jimmy-kimmel-live/show/34965">http://tv.yahoo.com/listings/jimmy-kimmel-live/show/34965</a>
            </LINK>
            <SHOWNAME> Jimmy Kimmel Live </SHOWNAME>
        </SHOW>
    </CHANNEL> .....
</CHANNELS>
</TVLISTINGSRESULTS>

```

The XML file is separated into channels. Each channel has a name and the XML file will contain a sequence of shows. Each show has a time slot, a link and a name. Note that if there is no link provided for a show you should just write "N/A", otherwise you should write the URL between the <LINK> tags.

- Your server-side script program should create the XML, save it into a file and

make the link titled “Yahoo! TV listings(XML)” (see Figure 2) point to it. Once the user clicks this link, the XML file should be shown in the window.

2.3 If the user clicks the hyperlink titled “Search...” (See figure 2) a new web page is displayed, containing 2 drop-down controls, named, “Select Channel” and “Select Time Slot”, as shown in Figure 4.

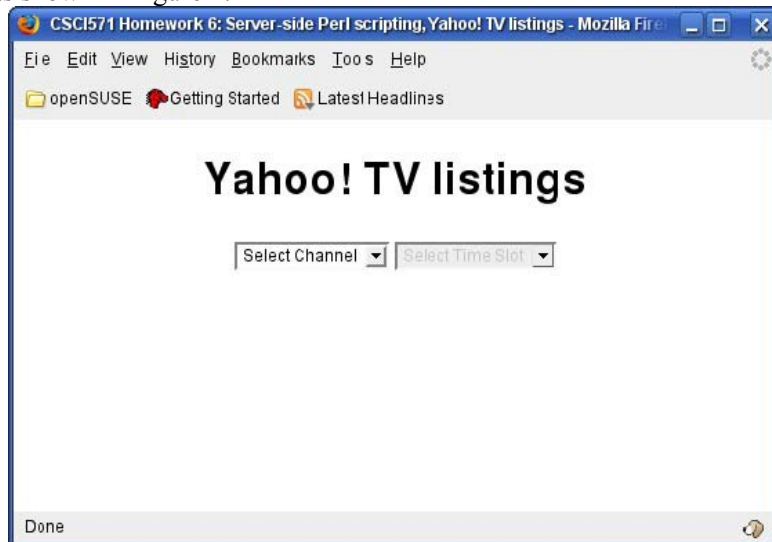


Figure 4 : Initially only the first dropdown is enabled and contains all the channel names. All other dropdowns are disabled. Once the Channel is selected, the next dropdown is enabled and is filled with the appropriate time slots.

Figure 5 shows the result of selecting “ABC”, and “3:00pm,-4:00pm” as of 17 Sept. 2008.

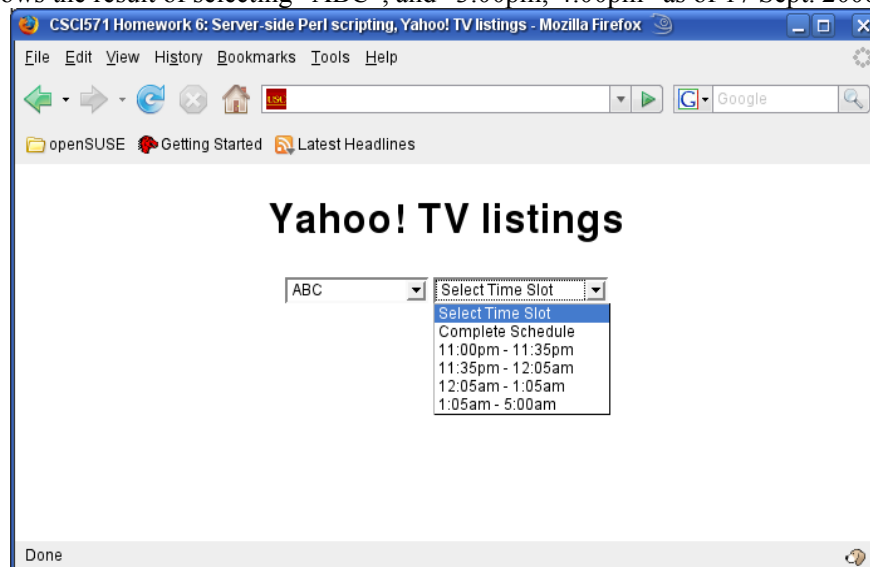


Figure 5

2.4 If the user selects one of the time slots, information about the show on the selected time-slot should be displayed on the **same** page without refreshing, as shown in the example in Figure 6. **Note** that the last line “More info.” is also a hyperlink to a valid URL. If the show has no URL (e.g. N/A in the XML <LINK> tag) you should display “None..” instead.

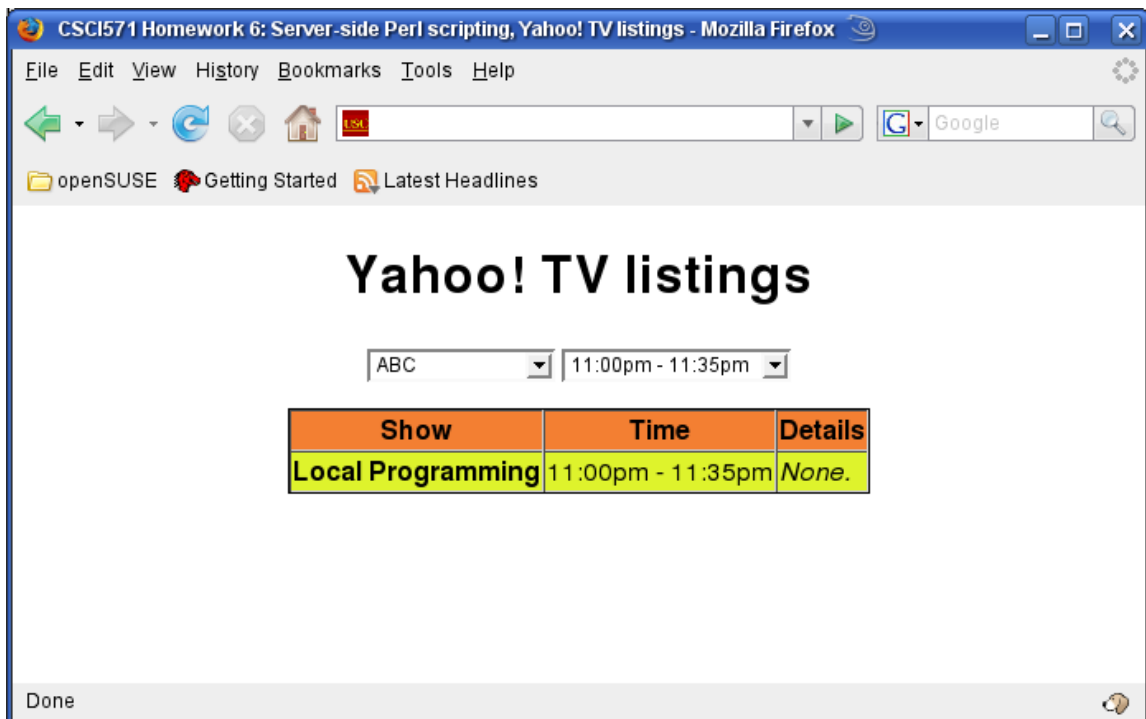
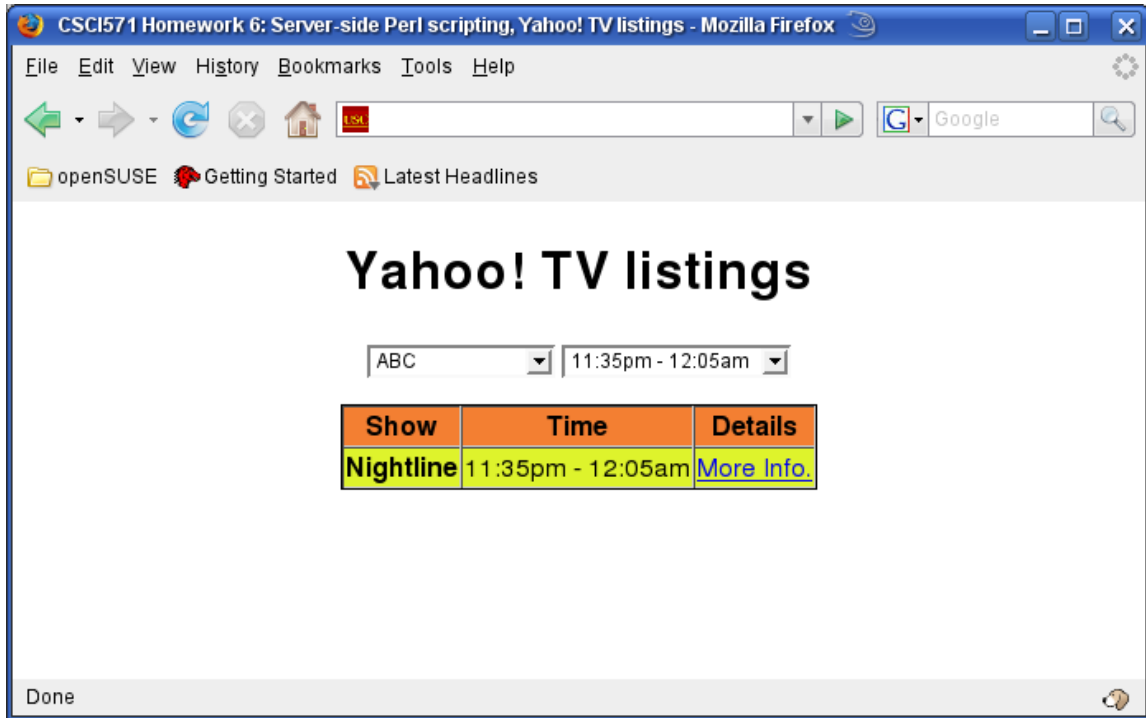


Figure 6: Please notice that there is NO submit button following the selection lists. That means you need to write some JavaScript to monitor the mouse events on the 2 drop-down selection lists. The second selection has no link to more information.

In addition to the different timeslots, the second dropdown menu should have “Complete Schedule” as the first entry. When selected all shows should be displayed in the same table as shown in the two examples below in Figure 7.

The screenshot shows a browser window with the title 'CSCI571 Homework 6: Server-side Perl scripting, Yahoo! TV listings - Mozilla Firefox'. The address bar contains 'usa'. The page content includes the heading 'Yahoo! TV listings', two dropdown menus (one for 'ABC' and one for 'Complete Schedule'), and a table with the following data:

Show	Time	Details
Local Programming	11:00pm - 11:35pm	None.
Nightline	11:35pm - 12:05am	More Info.
Jimmy Kimmel Live	12:05am - 1:05am	More Info.
Local Programming	1:05am - 5:00am	None.

The screenshot shows a browser window with the title 'CSCI571 Homework 6: Server-side Perl scripting, Yahoo! TV listings - Mozilla Firefox'. The address bar contains 'usa'. The page content includes the heading 'Yahoo! TV listings', two dropdown menus (one for 'CBS' and one for 'Complete Schedule'), and a table with the following data:

Show	Time	Details
Local Programming	11:00pm - 11:35pm	None.
Late Show With David Letterman	11:35pm - 12:35am	More Info.
Late Late Show with Craig Ferguson	12:35am - 1:35am	More Info.
Local Programming	1:35am - 6:00am	None.

Figure 7: Note that the time, show and details attributes appear in the first row of this table as opposed to the first row for individual selections.

3. Additional Requirements

- The data should be current. Graders will click this link and verify that your information is correct and identical to the current Yahoo! TV listings website at the time of their test.
- You should **NOT** design your own XML structure and tags, but follow the specification as given in section 2.3. A DTD is not required.
- You should **NOT** use any XML or HTML parsers. You should write your own parsing code in Perl.
- The code should work in both IE 6.0 and Firefox 2.0..
- Make sure your Apache web server is running so that the graders can do the grading.

4. Hints

- To retrieve a web page, you can use the Perl LWP module.

```
if(eval{require LWP::Simple;}){  
  
}else{  
    print "You need to install the Perl LWP module!<br>";  
    exit;  
}  
$url = "http://XXXXXX";  
$PageSource = LWP::Simple::get($url);
```

5. Files to Submit

Submit your HTML/JavaScript, Perl scripts, sample XML files electronically to the csci571 account so that they can be graded and compared to all other students' code via the MOSS code comparison tool. Also, you need to submit a README file that briefly describes the details of the organization of your program and how it works.