

## WEB-BASED SURVEYS: PROGRAMMING, METHODOLOGY, STATISTICAL SIGNIFICANCE

**Jan E. Heckman**

Librarian, Marine Sciences Liaison  
University of Connecticut  
Avery Point Campus - Library  
1084 Shennecossett Road  
Groton, CT 06340-6097  
860-405-9146  
libadm22@uconnvm.uconn.edu

**ABSTRACT:** The increased use of the World Wide Web (Web, or The Web) has caused a growing interest in using the Web for academic purposes. Delivery of information has been the major focus, with interactive instructional applications growing in importance.

The Web is also being used for information gathering. A web-based survey may increase response numbers because of ease of use, and the data gathered can be immediately used as the data file for a statistical package. Costs in both time and money may be reduced while accuracy could be increased.

The methodological and statistical rigors of academic surveys, as they relate to Web-based surveys, have yet to be fully investigated. This presentation will address those issues while giving examples of both method and use.

### INTRODUCTION

In 1972 the results of a survey done by Ellis Mount was published wherein it was noted that 50% of science/engineering librarians possessed science degrees. (Mount 1972) There were a total of three questions about "background and training". The 50% science degree figure has been used extensively in numerous articles on the subject of science/engineering libraries over the years. The study in 1972 surveyed 22 large academic universities and like most surveys done in the 70's was time consuming and limited.

The survey I will describe, mine, is meant to poll those very same 22 universities used in the Mount survey of 1972 but greatly expand on it. My present day survey will focus on questions asked about "background and training" plus survey a much larger number of science librarians/information specialists. With the use of the World Wide Web (Web), Hypertext Markup Language (HTML) forms, and Common Gateway Interface (CGI) script, more librarians can be surveyed with less effort and increased accuracy.

The methodology of my survey, e.g. a Web-based survey with automatic tabulation of data in a format that can be uploaded into a statistical software package is relatively new. The idea of a survey which doesn't have to be printed and sent to the target population, that is easy to fill out and will produce not only a data file but a count of respondents is the gold at the end of the rainbow. Questions still exist as to how to control a Web-based survey given the open world of the Internet and The Web. A literature search did not provide many answers. It can be said that the articles I did find thought Web-based surveys would indeed be easier and more accurate.

Statistical soundness of Web-based surveys is a larger issue. There is an inherent bias in any survey that relies on Web access, but in the case of science librarians I don't see this as significant. My survey will contact not only the science librarians at the 22 universities that Ellis Mount surveyed and control for that group, but try to reach out to a much larger science librarian community that is also international in scope. The larger community, which can be reached via electronic discussion lists, is not a well-defined group, but I will attempt to define them as they request access to the survey. This attempt will likely produce a highly critical response from survey statisticians, but science librarians are in many cases a self-defined group and building the survey population is one of the tasks of any survey. Akin to "Who's Who in ..." if you profess to be a science librarian, odds are, that you are.

There will be three somewhat distinct groups contacted. One group is the original 22 universities and then I will choose another set of schools that are smaller. In both cases I intend to correspond with an individual in the library and ask for a count of full and part-time science librarians in their given institution. I will then send them that number of cards with a username and password on each. The third group will be individuals who respond to messages sent out over science librarian related electronic discussion group lists. Those who respond will be sent a username and password. There will be three different usernames, one for each group. When a survey is finished and submitted the CGI script will increment a counter in order to track response rates for each group. Passwords will be distinct for each individual and will mainly be used to stop anyone not contacted from accessing the survey. When individuals access and submit a survey their passwords will be automatically removed from a list of valid passwords. I can't imagine someone wanted to fill out the survey more than once, unless he made a mistake and wished to correct it, nevertheless it won't be possible. The file of passwords for the third group will have to be updated on a regular basis as more people request a username and password. A copy will have to be accessed from the server, names added, and quickly put back.

#### **METHODOLOGY-THE CODING: HTML forms and data formatting, CGI SCRIPT**

The Web allows for a user to input data or requests which will be recorded or activate responses by other computers. An increasingly common Web page is one with buttons or boxes to choose or text areas to fill out. This type of page is obviously conducive to

surveys. The HTML forms which produce a button to choose, a pull down menu of choices, a checkbox to choose, and a text area are illustrated.

Most HTML has opening and closing codes (called tags) to generate different formatting or capabilities of a Web page. Forms must start with <form> and end with a closing </form> tags. *HTML coding is identified with Italics and proper HTML tags. Other examples will not have full HTML coding only the forms coding*  
Examples:

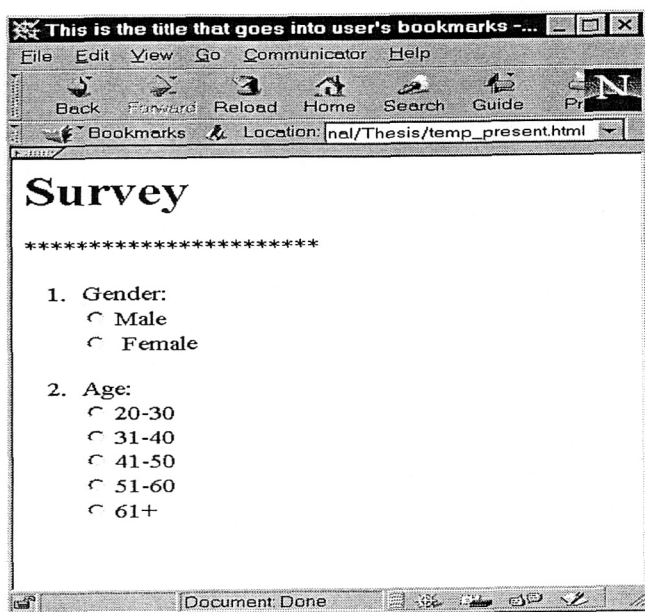
```
<!-- This is a comment and will not be displayed -->
```

### “Radio” Button Coding

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 3.2 Final//EN">
<HTML>
<HEAD>
<TITLE>This is the title that goes into user's bookmarks</TITLE>
</HEAD>
<BODY>
<form action="cgi/bin/mysurvey" method="POST" >
<h1>Survey</h1>
*****
<ol> <!--this is the start of an ordered list -->
<li> <!-- list item --> Gender:<br>
<INPUT TYPE="radio" NAME="gender" VALUE="1">Male<br>
<INPUT TYPE="radio" NAME="gender" VALUE="2">Female<p>
<li> Age:<br>
<INPUT TYPE="radio" NAME="age" VALUE="1">20-30<br>
<INPUT TYPE="radio" NAME="age" VALUE="2">31-40<br>
<INPUT TYPE="radio" NAME="age" VALUE="3">41-50<br>
<INPUT TYPE="radio" NAME="age" VALUE="4">51-60<br>
<INPUT TYPE="radio" NAME="age" VALUE="5">61+<br><p>
</ol> <!-- Close of ordered list -->
</form><!-- Closing form tag. This would normally be at the end of survey form -->
</BODY>
</HTML>
```

With the INPUT TYPE="radio" only one button can be chosen at a time with the group which is defined by NAME="gender". This NAME is also the variable name, which will be returned to a CGI script, which will encode the VALUE (1 or 2 for variable "gender") in the data file used in a statistical software package.

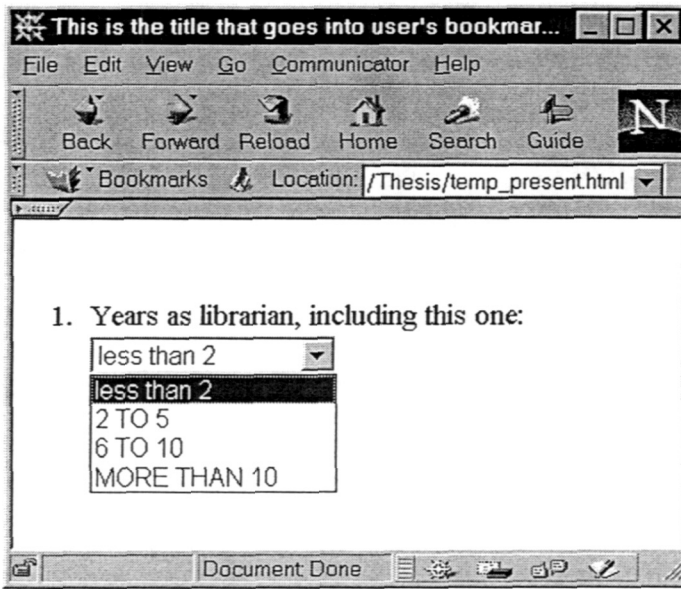
The HTML code above produces the following:



**Pull down menu code:**

```
<i>Years as librarian, including this one:<br>
<SELECT size=1 name="years">
<OPTION VALUE="1">less than 2
<OPTION VALUE="2">2 TO 5
<OPTION VALUE="3">6 TO 10
<OPTION VALUE="4">MORE THAN 10
</SELECT>
```

The HTML code produces the following:

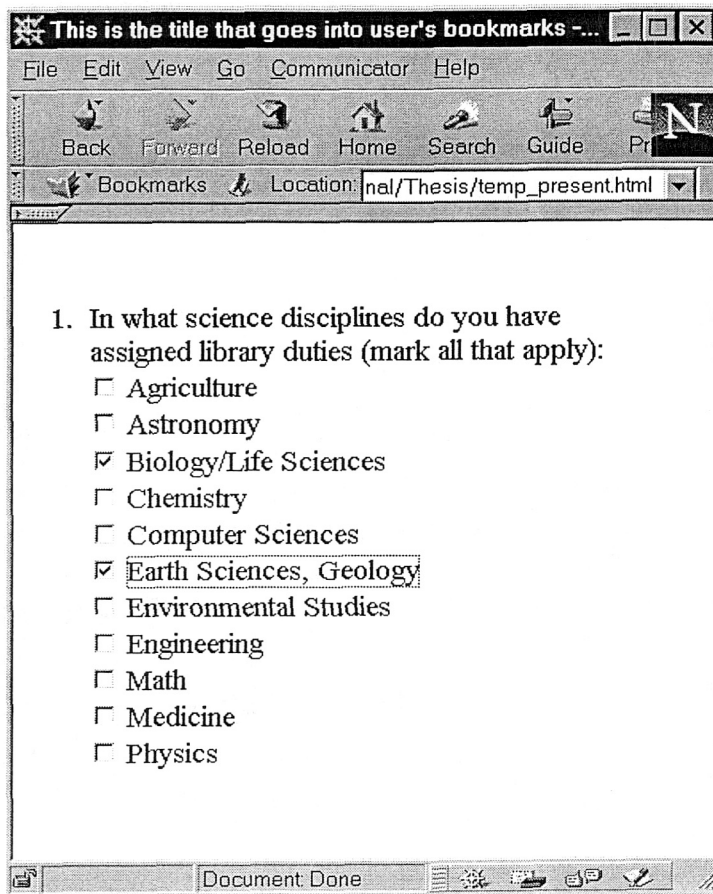


**Checkbox code-** Note that each checkbox has a unique variable NAME because more than one checkbox can be chosen.

*<i>*In what science disciplines do you have assigned library duties (mark all that apply):*<br>*

```
<INPUT type="checkbox" name="agri_disc" value="1">Agriculture<br>
<INPUT type="checkbox" name="astron_disc" value="2">Astronomy<br>
<INPUT type="checkbox" name="bio_disc" value="3">Biology/Life Sciences<br>
<INPUT type="checkbox" name="chem_disc" value="4">Chemistry<br>
<INPUT type="checkbox" name="comp_disc" value="5">Computer Sciences<br>
<INPUT type="checkbox" name="earth_disc" value="6">Earth Sciences, Geology<br>
<INPUT type="checkbox" name="environ_disc" value="7">Environmental Studies<br>
<INPUT type="checkbox" name="engin_disc" value="8">Engineering<br>
<INPUT type="checkbox" name="math_disc" value="9">Math<br>
<INPUT type="checkbox" name="med_disc" value="10">Medicine<br>
<INPUT type="checkbox" name="physic_disc" value="11">Physics<br>
<p>
```

The HTML code above produces the following:



Text area code

Other:

```
<INPUT type="text" NAME="hired" maxlength="75" size=30> <p>
```

The HTML code above produces the following:



### Username and Password coding:

```
<form method="post" action="http://survey.html">
```

Username:

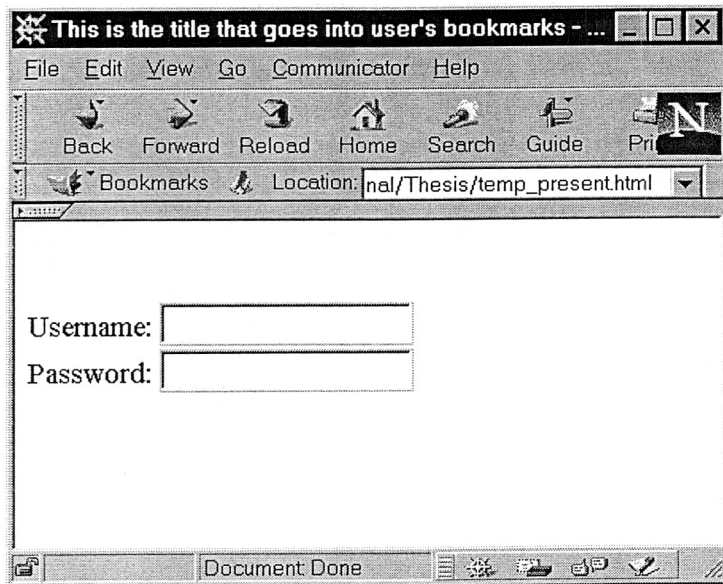
```
<INPUT TYPE="text" NAME="username" SIZE=15><br>
```

Password:

```
<INPUT TYPE="password" NAME="password" SIZE=15>
```

```
</form>
```

The HTML coding produces the following:



#### DISCUSSION OF HTML CODING :

The HTML coding of forms is relatively straightforward and uses the simple syntax illustrated above. Special attention should be paid to keeping the variable names straight and the values to be returned unique so the CGI script can place the correct return value in the proper place in a data file. Radio button and drop down menu groups can have one variable name but each choice within the group needs a unique return value. However, more than one checkbox can be chosen at a time, thus each checkbox needs a unique variable name. Another coding issue before CGI script is written is allowing for non-response both by failing to mark or choosing not to mark. It may not matter statistically because in both cases the individual's answers would have to be thrown out, at least for that question, however one may want to track the two types of non-response. If a question continues to be skipped by conscious choice the question itself may have to be changed or dropped. This method would also allow for testing of a survey.

A method for handling both type of non-responses is to have "Choose not to answer" in bright red text as one of the options available. The other "failed to mark" situation can be handled by the CGI script returning a null value.



## DISCUSSION OF CGI (COMMON GATEWAY INTERFACE) SCRIPT:

Common Gateway Interface (CGI) is a “protocol that describes the standard method of communications between a web server and an external software applications” (Nader 1998). The protocol should not be confused with a “script” which can be written in a variety of programming languages. My survey will use PERL (Practical Extraction and Reporting Language).

This sub process handles the input from the HTML forms:

```
-----  
sub UnWeb {  
  
# Get the input  
read(STDIN, $buffer, $ENV{'CONTENT_LENGTH'});  
  
# Split the name-value pairs  
@pairs = split(/&/, $buffer);  
  
foreach $pair (@pairs) {  
($name, $value) = split(/=/, $pair);  
  
# Decode  
$value =~ tr/+// ;  
$value =~ s/%([a-fA-F0-9][a-fA-F0-9])/pack("C", hex($1))/eg;  
$value =~ s/<!--(.|\n)*-->//g;  
  
if ($allow_html != 1) {  
$value =~ s/<([^\>])\n*>//g;  
}  
else {  
unless ($name eq 'body') {  
$value =~ s/<([^\>])\n*>//g;  
}  
}  
  
$in{$name} = $value;  
}  
  
}
```

## CONCLUSION:

The ability to use the increasing reach of the Internet and World Wide Web, automatic tabulation of survey response data and the ease of filling out an online survey combined

may dramatically change the effectiveness of surveys. The issue of how populations are defined, contacted, and polled while maintaining statistical rigor is an issue as yet unresolved. I'm hoping my survey is a step in the right direction.

## REFERENCES

- Baker, Reginald P. 1998. The CASIC future. In: *Computer assisted survey information collection* (ed.by Mick P. Couper ... [et al.]) Chap. 29. Wiley, NY.
- Clayton, Richard L. 1998. Business surveys of the future: the World Wide Web as a data collection methodology. In: *Computer assisted survey information collection* (ed.by Mick P. Couper ... [et al.]) Chap. 27. Wiley, NY.
- Mount, Ellis 1972. University science and engineering libraries - a survey. (ERIC doc. no.) ED 068 108
- Nader, Jonar C. 1998. *Prentice Hall's Illustrated Dictionary of Computing*. Prentice Hall, Sydney.
- Ramos, Magdalena & Sedivi, Barbara M. & Sweet, Elizabeth M. 1998. Computerized self-administered questionnaires. In: *Computer assisted survey information collection* (ed.by Mick P. Couper ... [et al.]) Chap. 20. Wiley, NY.

## WEB RESOURCES

- Rutter, Richard. "Sizzling jalfrezi - HTML by example." [Online.] Available: <http://vzone.virgin.net/sizzling.jalfrezi/iniframe.htm> [Sept.1999].