

# EECS 4313

## Software Engineering Testing



### Topic 12:

### Website Testing

**Zhen Ming (Jack) Jiang**

# How to test a website

- Easiest way to start is by treating the web site as a black box.
  - Look at a sample website such as [www.amazon.com](http://www.amazon.com) to get a sense of the scale of such an endeavor.
  - Treat each page as a state with hyperlinks as state transitions.

# Website text

- Web page text should be treated like documentation and be tested using the techniques we described previously.
- Check for ...
  - correctness of contact information e.g., phone numbers, addresses
  - correctness of dates and copyright notices
  - title bar text, bookmark text on browser's favorites
  - correctness of the dynamic contents (e.g., mouse over text)
  - layout issues when browser window is resized

# Website hyperlinks

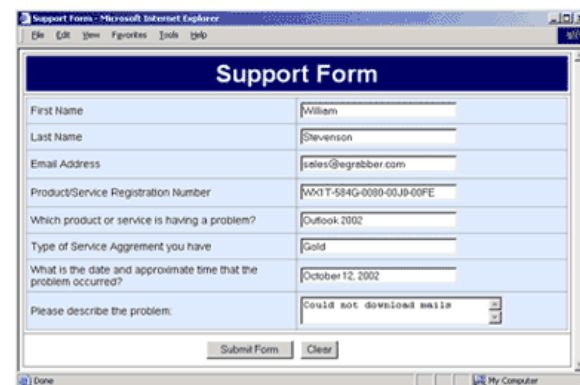
- Each link should be checked to make sure it jumps to the correct destination or website.
- Ensure that hyperlinks are obvious:
  - E.g., underlined text, mouse pointer changes
- If the link opens an e-mail message, test it
  - Send an e-mail and verify that you get a response.
- Check for orphan pages that are part of the website but cannot be accessed through a hyperlink.
  - Someone forgot to create the link
  - Might be intentional ... Google may find it, though

# Website graphics

- Do all graphics load and display properly?
- Is a graphic missing or incorrectly named?
- Does the website intermix text and graphics?
  - Does the text wrap around the graphics?
  - What happens when the browser window is re-sized?
- Does the page load fast enough?
  - Are there too many graphics?
  - Did you try to test the website on a slow connection (e.g., dialup connection) instead of a high-speed connection (e.g., cable or LAN)?

# Website forms

- Forms are the text boxes, list boxes, and other fields for entering and selecting information on the web page.
- Are the form fields positioned properly?
- Are the fields the correct size?
- Do they accept correct data?
- Do they reject bad data?
- Are optional fields really optional?
- A favorite entry point for buffer overflow attacks.



The screenshot shows a web browser window titled "Support Form - Microsoft Internet Explorer". The form itself is titled "Support Form" and contains the following fields:

First Name	William
Last Name	Devenson
Email Address	jsales@regrabber.com
Product/Service Registration Number	WWT-584G-0080-00JD-00FE
Which product or service is having a problem?	Outlook 2002
Type of Service Agreement you have	Gold
What is the date and approximate time that the problem occurred?	October 12, 2002
Please describe the problem:	Could not download mail

At the bottom of the form, there are two buttons: "Submit Form" and "Clear".

# “Grey-box” testing

- A mixture of white-box and black-box testing.
- You stick to black-box testing primarily and supplement it by taking a peek at the HTML to figure out how the website works.
- For website testing it is worth looking at the HTML code.
  - It’s there, easy to look at, why not?
- Not looking at the HTML code is wasteful, especially since HTML is such a simple language.
  - HTML is a tagging language for text and graphics.
  - To create dynamic web content requires that HTML be supplemented by programming code (e.g., Java applets, ActiveX, Javascript, VBScript, CGI, Perl).



# An overview of HTML



# HTML

Final Report

European Conference on Expert Systems

*boldface*

Submitted by Justin Parker

*Center*

First of all, our thanks go out to the following sponsors for their support of the conference and its supplemental activities.

Allied Interactive

Sybernetics, Inc.

Dynamic Solutions of New Jersey

*make these bullets*

The conference was a great success. It ran a full four days, including workshops and special sessions. Subjective feedback from conference attendees was largely positive, and financially the revenues resulted in a surplus of over \$10,000.

Figure 16.2

A marked-up document

# HTML

```
<HTML>
  <HEAD>
    <TITLE>Student Dynamics</TITLE>
  </HEAD>
  <BODY>
    <CENTER><IMG SRC="stuDynamics.gif"></CENTER>
    <HR>
    <CENTER><I>A student-based community organization.</I></CENTER>
    <HR>
    <P>Plan to participate in our upcoming events:</P>
    <UL>
      <LI>Visit Children's Hospital (March 23)</LI>
      <LI>Help clean renovated school grounds (April 4)</LI>
      <LI>Continue the <a href="outreach.html">K-12 outreach
        program.</a> <IMG SRC="updated.gif"></LI>
    </UL>
    <P><B><I>You should be involved!</B></I> Help our active and
      energetic team make a difference in the lives of people. Our
      <a href="execBoard.html">executive board</a> is always willing
      to answer any questions you may have.</P>
    <P>How can we help you? What suggestions do you have regarding
      new activities? <a href="suggestions.html">Let us know!</a></P>
    <P>We are always in need of donations: equipment, money, or
      (better yet) you!</P>
    <CENTER><H3>Join us for our Spring Picnic in May!</H3></CENTER>
  </BODY>
</HTML>
```

Figure 16.4  
The HTML document  
defining the Student  
Dynamics Web page

# HTML

- Tags are enclosed in angle brackets (< . . . >)
- Words such as HEAD, TITLE, and BODY are called elements and specify the type of the tag
- Tags are often used in pairs, with a start tag such as <BODY> and a corresponding end tag with a / before the element name, such as </BODY>

# HTML

- The browser determines how the page should be displayed based on the tags
- The browser
  - Ignores the way we format the HTML document using carriage returns, extra spaces, and blank lines
  - Takes into account the width and height of the browser window
  - Reformats the contents to fit your browser window

# Basic HTML Formatting

- The paragraph tags (`<P> . . . </P>`) specify text that should be treated as a separate paragraph
- The center tags (`<CENTER> . . . </CENTER>`) indicate that the enclosed information should be centered in the browser window

# Basic HTML Formatting

- The B, I, and U elements are used to indicate that the enclosed text should be bold, italic, or underlined, respectively
- The <HR> tag inserts a horizontal rule (that is, a line) across the page

# Basic HTML Formatting

- We often have cause to display a list of items

The UL element stands for an unordered list, and the LI element represents a list item

- Several elements are used to define headings in a document

There are six predefined heading elements defined in HTML: H1, H2, H3, H4, H5, and H6

# Images and Links

- Many tags can contain attributes that indicate additional details about the information or how the enclosed information should be displayed
  - An image can be incorporated into a web page using the IMG element, which takes an attribute that identifies the image file to display
    - `<IMG SRC = "myPicture.gif">`
- A link is specified using the element A, which stands for anchor
- The tag includes an attribute called HREF that specifies the URL of the destination document.
  - For example `<A HREF = "http://duke.csc.villanova.edu/docs/"> Documentation Central!</A>`



# White-box testing

- To make sure you find the important bugs you should have some knowledge of the website's programming:
  - Dynamic content
  - Database-driven web pages
  - Programmatically created web pages
  - Server performance and loading and security

# Dynamic content

- Dynamic content is graphics and text that changes based on certain conditions.
  - E.g., time of day, weather, stock tickers
- Client-side programming involve embedding scripting code (e.g., JavaScript) into HTML.
  - Less efficient if the computations are expensive
  - Code is local, easy to access for testing
- Server-side programming is code located on the server.
  - More efficient for intensive calculations
  - Requires access to the web server to view the code (might be a problem for testing).

# Database-driven web pages

- Most E-commerce websites that show catalogs or inventories are database driven.
  - E.g., Amazon.com, BestBuy.ca
- Data is pulled from the database, formatted into HTML and sent to the web browser for viewing. A three tier architecture is used:
  - Tier 1: web browsers (presentation)
  - Tier 2: web page formatter (converts data from Tier 3 to HTML)
  - Tier 3: database (queried by Tier 2)

# Programmatically - created web pages

- HTML or code is generated by a program.
  - E.g., a web designer may drag and drop elements in a layout program, press a button, and get an HTML page.
- Compilers do this all the time, except they translate high-level language code (Java source code) into low-level code that can be executed natively or on a VM (e.g., byte code).



# A short intro to Java Server Page (JSP)

# Interactive Web Pages

- When HTML was first developed, there was no way to interact with the information and pictures presented in a web page
- As users have clamored for a more dynamic web, new technologies were developed to accommodate these requests
- Many of the new ideas were offshoots of the newly developed Java programming language

# Java Applets

- **Java applet** A program that is designed to be embedded into an HTML document and transferred over the Web to someone who wants to run the program

An applet is embedded into an HTML document using the APPLET tag

```
<APPLET code="MyApplet.class" width=250  
height=150 ></APPLET>
```

Example: <http://polymer.bu.edu/java/java/winning/lotteryapplet.html>  
(Make sure change security setting in Java and run inside IE)

# Java Applets

- A browser has a built-in interpreter that executes the applet, allowing the user to interact with it.
  - Consider the difficulties inherent in this situation
  - How can we execute a program that was written on one type of computer on possibly many other types of computers?



# Java Applets

- Java programs are compiled into Bytecode, a low-level representation of a program that is not the machine code for any particular type of CPU
- Java applets are restricted as to what they can do
  - The Java language has a carefully constructed security model
  - An applet, for instance, cannot access any local files or change any system settings

# Java Server Pages

- A Java Server Page, or JSP, is a web page that has **JSP scriptlets** embedded in them
- **Scriptlet** A small piece of executable code intertwined among regular HTML content

# Java Server Pages

- A JSP scriptlet is encased in special tags beginning with `<%` and ending with `%>`
- Imagine JSP scriptlets as having the expressive power of a full programming language

```
<H3>  
<%  
out.println ("hello there");  
%>  
</H3>
```

# Java Server Pages

- Note that JSPs are executed on the server side where the web page resides
- By the time it arrives at your computer, all active processing has taken place, producing a static (though dynamically created) web page
- JSPs are particularly good for coordinating the interaction between a web page and an underlying database



A short demo on  
Dell DVD Store (DS2)

# Dell DVD Store

**Welcome to the DVD Store - Click below to begin shopping**

**Your previous purchases:**

Title	Actor	People who liked this DVD also liked
ACE WEST	CHARLIZE WINSLET	AIRPORT WAGON
ACADEMY GOLDFINGER	CHRIS RYAN	AFRICAN DOLLS
ACE WEST	CHARLIZE WINSLET	AIRPORT WAGON

[Start Shopping](#)

---

**Welcome to the DVD Store - Click below to begin shopping**

**Your previous purchases:**

Title	Actor	People who liked this DVD also liked
AFFAIR LOVERBOY	CHARLES KELLY	ACE NEIGHBORS
AIRPORT ROXANNE	WILLEM DAY	ACE FLINTSTONES
AFRICAN SPIRITED	KIRK NORTON	AFRICAN CONNECTION
AIRPLANE TREATMENT	ORLANDO MINELLI	AIRPLANE UNDEFEATED
AIRPLANE NIGHTMARE	WARREN FISHER	ALADDIN TYCOON

[Start Shopping](#)

---

# Server performance loading and security

- Popular websites can receive millions of hits per day.
  - [www.youtube.com](http://www.youtube.com) hits 100 Million Videos per day
  - Check out <http://www.youtube.com/watch?v=jP26ZYhQktY> for a software testing video!
- Each hit requires a download of data from the website's server to the browser's computer.
- You need to simulate millions of connections and downloads to test a system for performance and loading.
  - Stay tuned for the topic on load testing!
- Security issues include:
  - Denial of service (DoS) attacks
  - Buffer overflow attacks

# Configuration testing and compatibility testing

- Hardware platform
  - Mac, PC, smartphone?
- Browser software version
  - Firefox, IE, Chrome, Safari
- Browser plug-ins
  - To play specific types of audio or video files
- Browser options
  - Security options, ALT text, plug-in, pop ups
- Video resolution and color depth
  - 640x480, 800x600, 1024x768, 1280x1024, 256 colors, 16 colors
- Text size
  - Small fonts, medium fonts, large fonts
- Connection speed
  - DSL, modems of varying speed.



# Usability testing:

## Jacob Nielsen's top 10 mistakes in web design

1. Gratuitous use of bleeding-edge technology
2. Scrolling text, marquees, and constantly running animations
3. Long scrolling pages
4. Non-standard link colors
5. Outdated information (need website maintenance)
6. Overly long download times (less than 10 sec to load)
7. Lack of navigation support
8. Orphan pages
9. Complex website addresses (URLs)
10. Using Frames (just open another window ...)

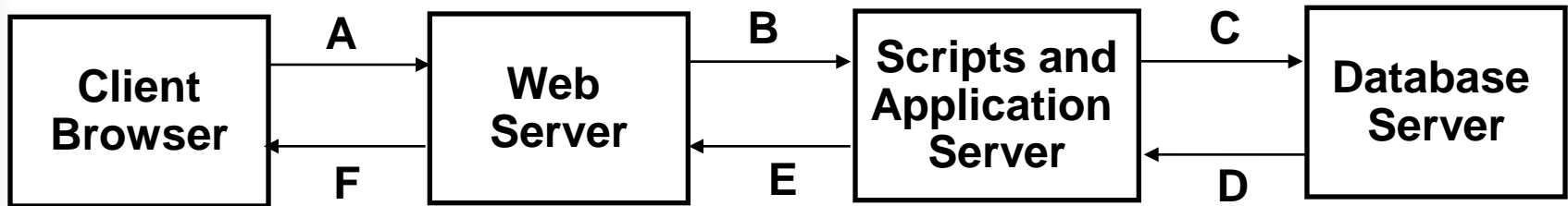
# Website testing tools

- Website testing is very labor-intensive.
- Tools that automatically check websites for:
  - Browser compatibility
  - Performance problems
  - Broken hyperlinks
  - HTML standard adherence
  - Spelling on text
- Web browser automation testing
  - <http://www.seleniumhq.org/>

# Web applications

- Website - requests are navigational requests.
- Web Application - requests can affect the state of the business logic on the server.

# A web application



**A: Request**

**B: Request for page generation**

**C: SQL command**

**D: Result set**

**E: Dynamically generated page**

**F: Response (dynamic | static page)**

# Web application failures

A study by the Business Internet Group of San Francisco found that:

- 72% (29/40) leading e-commerce sites, and
- 68% (28/41) government sites contained Web application failures
  - 25 “technical errors”
    - E.g., page not found, multiple attempts to subscribe to a service
  - 3 data errors
    - E.g., page without text, wrong page returned

# What type of bugs?

## - An Empirical Study by Kallepalli and Tian

- Analyzed the logs of the Web pages of the School of Engineering and Applied Science at Southern Methodist University for usage and failure information.
- Looked at data from 26 consecutive days.
- Total number of “hits” 762,971 (~30,000 hits/day)
  - A “hit” is registered for each Web page if one of the following happens:
    - The HTML file corresponding to a page is requested.
    - Any graphics within the HTML page is requested.

# Bug types

- Permission denied
- No such file or directory
- Stale NFS handle
- Client denied by server configuration
- File does not exist
- Invalid method in request
- Invalid URL in request connection
- Mod\_mime\_magic
- Request failed
- Script not found or unable to start
- Connection reset by peer

# Distribution of bugs

Permission denied	2079
No such file or directory	14
Stale NFS handle	4
Client denied by server configuration	2
File does not exist	28631
Invalid method in request	0
Invalid URL in request connection	1
Mod_mime_magic	1
Request failed	1
Script not found or unable to start	27
Connection reset by peer	0