THE BCS PROFESSIONAL EXAMINATION
Professional Graduate Diploma

April 2007

EXAMINERS' REPORT

Web Engineering

General Comments

This paper was offered for the first time. It is intended to supersede the paper *World Wide Web: Beyond the Basics* and it is not possible to take both examinations concurrently. With the exception of question 3, the quality of answers to this paper were very good and candidates appear well-prepared and did not typically suffer to the same extent from the weaknesses as displayed previously in WWW:BTB (verbose answers, textbook theory).

Question 1

a)

i) Explain the differences between HTML 4.01 and XML. **(2 marks)**

ii) Explain, with reference to XML, the purpose of the Document Type Definition and the Document Type Declaration. **(2 marks)**

iii) What is XSLT and what is it used for? **(3 marks)**

b) Referring to the XML document in Figure 1, write a DTD which **enforces** the following NINE constraints:

sequence of elements is as shown in the XML code above (i.e. `library` is a container of `book` elements)
`book` element must be present zero or more times
The two attributes in `book` are mandatory
Exactly one `title` element must be present
Exactly one `author` element must be present
`author_email` element is optional
`book_url` element is optional
`book_url` is an `EMPTY` element
At least one `review` element must be present **(9 marks)**

2 The website designers wish the XML list of books to be displayed as a web page as indicated in Figure 2. Using the HTML template provided in Figure 3, complete the missing code (the section marked `<!–TO BE COMPLETED -->`) to accomplish this. **(9 marks)**

Answer pointers

a)

i)

HTML (Hypertext mark-up language) is a mark-up language designed for the creation of web pages. HTML 4.01 is based on SGML.
XML (Extensible mark-up language is a general purpose language, and is a way of describing data. XML allows for user-defined tags, and is both human and machine readable at the same time.

ii)

The Document Type Definition (DTD) is a schema defines the syntax of mark-up constructs (elements and attributes). It allows for checking whether a particular XML document using this schema is valid.

The Document Type Declaration (DOCTYPE) is the statement at the top of an XML file that specifies the location of the Document Type Definition to use.

iii)

XSLT (XML Transformations) is a language for transforming XML documents into other XML documents (e.g. converting XML into XHTML for displaying in a web browser). It works on valid XML, viewing it as a source tree of nodes (elements) which are converted to a result tree, using rules.

b)

```
<!DOCTYPE Library [
  <!ELEMENT Library (book*)>
  <!ELEMENT book (title,
        author,author_email?,book_url?,publisher,review+)>
  <!ATTLIST book isbn CDATA #REQUIRED edition CDATA #REQUIRED>
  <!ELEMENT title (#PCDATA)>
  <!ELEMENT author (#PCDATA)>
  <!ELEMENT author_email (#PCDATA)>
  <!ELEMENT book_url EMPTY>
  <!ATTLIST book_url page CDATA #REQUIRED>
  <!ELEMENT publisher (#PCDATA)>
  <!ELEMENT review (#PCDATA)>
]>
```

c)

```
<xsl:for-each select="Library/book">
<tr>
  <td><xsl:value-of select="@isbn"/></td>
  <td><xsl:value-of select="title"/></td>
  <td><xsl:value-of select="author"/></td>
  <td><xsl:value-of select="publisher "/></td>
  <td>
  <xsl:choose>
    <xsl:when test="book_url/@page" >
      <xsl:value-of select="book_url/@page "/>
    </xsl:when>
    <xsl:otherwise>
      No website for the book
    </xsl:otherwise>
  </xsl:choose>
  </td>
</tr>
</xsl:for-each>
```

Examiners' Guidance Notes

Two-thirds of the candidates attempted this question and most of the candidates were well prepared and as such the quality of the answers was very good. There were no specific systematic mistakes.

It should be noted that the scope of questions on XML in future years may not be limited to just this specific style of question, and a broad coverage of XML should be maintained by candidates preparing in the future.

Question 2

a) Write HTML to construct the form as indicated in Figure 4. When the button marked "Log in" is pressed, the form should submit to a script called `login.php` **(3 marks)**

b)

i) Why is it important to validate data? **(2 marks)**

ii) Describe, with an example for each, when data should be validated at:

- The client end
- The server end **(3 marks)**

c) Write code (either at the client side in JavaScript, or at the server side in ASP, PHP, or Perl) to validate that:

i) The name field exists, and consists of alphabetic characters or spaces only. **(4 marks)**

ii) The password field has at least 6 characters, and contains at least one non-alphanumeric character. **(4 marks)**

The code should return an appropriate and informative error message if the validation fails.

d) The server has a database named **`Security`**, with a single table named `User` (as shown in Figure 5). You may assume the web server and the database server are the same machine.

For parts ii) and iii) below, assume that the php script `login.php` has already established a valid connection to this database. In all cases, you should state the language you are using (ASP, PHP or Perl).

i) Write code to connect to the database. **(1 mark)**

ii) Write code to store the details submitted from the form into the database. **(3 marks)**

iii) Write code to retrieve and display (in an appropriate format) the details of *all* usernames entered with the password "`password!`" **(5 marks)**

Note: The following SQL syntax may be useful to accomplish these tasks:

- *`INSERT INTO tbl_name (col1, col2, …) VALUES (val1, val2, …);`*
- *`SELECT * FROM tbl_name WHERE col1 = val1;`*

(Where `tbl_name`, `col1`, `val1` etc. are to be replaced with appropriate values)

Answer pointers

a)

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
"http://www.w3.org/TR/html4/loose.dtd">
  <head>
    <title>Login Screen</title>
    <meta http-equiv="Content-Type" content="text/html; charset=iso-8859-
        1">
  </head>
  <body>
  <h1>Login Screen</h1>
    <form action="password.php" method="get">
      <p>Enter your name: <input type="text" value="" name="name">
      <br>Enter your password: <input type="password" value="1234"
        name="password">
      <br><input type="submit" value="Log in">
      <input type="reset" value="Clear Form">
      </p>
    </form>
  </body>
</html>
```

b)

i) For a program to operate effectively, correct data is vital. Data validation is important because it is not possible to predict what the user agent will enter into input fields. Data may be corrupt, missing, badly formed or even maliciously designed to exploit weaknesses in the program.

ii)

Data should be validated at client end when it is important to have quick feedback to the user of any errors e.g. forgetting to fill in a postcode.

Whilst it is true that all important data should be validated at server end (regardless of the presence of client-side validation), and this should be given partial credit, an answer for full marks should consider a case where client side validation is not feasible, and validation must to take place at the server end. E.g. if it relies on checking the data using some external means (e.g. a Postcode against the Royal Mail database).

c)

Note: all examples for c) and d) are given in PHP, but ASP or Perl are also acceptable.

i)

```
// where 'name' is the name of the textfield defined in part a)
$username = $_GET['name'];
$pattern = '/^[\w ]+$/';
$matches = "";
if(!preg_match($pattern, $username))
{
 echo "<p>User name must contain only alphabetic letters and
        spaces.</p>\n";
} else {
 // SOME CODE
}
```

ii)

```
// where 'name' is the name of the textfield defined in part a)
$password = $_GET['password'];
$pattern = '/[^a-zA-Z0-9 ]$/';
$matches = "";
if(!(preg_match($pattern, $password) && (strlen($username) >= 6)))
{
 echo "<p>User name must be at least 6 characters in length, and
        contain at least one non alphanumeric character (e.g.
        '!').</p>\n";
} else {
 // SOME CODE
}
```

d)

i)

```
$link = mysql_connect('localhost:3307', 'user', 'password');
```

(Note: mysql_connect and server name localhost or 127.0.0.1 are required)

ii)

```
// assume connection already established from part d) i)
$username = $_GET['name'];
$password = $_GET['password'];
$username = addslashes($username);
$password = addslashes($password);
$result = mysql_select_db("SECURITY", $link);
if (! $result) {
   echo "Failed to connect to database.\n";
}
else
{
  $result = mysql_query("insert into User (username, password) values
        ('$username', '$password');", $link);

  if (! $result) {
     echo "<p>MySQL Error: " . mysql_error($link) . "</p>\n";
  }
  else
  {
    echo "<p>User " . $username . " added to the database.</p>";
  }
}
mysql_close($link);
```

iii)

```
// assume connection already established from part d) i)
$key = "password!";
$result = mysql_select_db("SECURITY", $link);
if (! $result) {
   echo "Failed to connect to database.\n";
}
else
{
  $result = mysql_query("select Username from User where Password =
        '$key'", $link);
  if (! $result) {
     echo "<p>MySQL Error: " . mysql_error($link) . "</p>\n";
  }
  else
  {
    $all_results = mysql_fetch_array($result);
    if(empty($all_results))
    {
      echo "<p>No users have the password: " . $key . "</p>\n";
    }
    else
    {
      echo "<table>\n";
      echo "  <tr><th>Username</th></tr>\n";
      foreach($all_results as $username)
      {
        echo "  <tr><td>$username</td></tr>\n";
      }
      echo "</table>\n";
  }
}
mysql_close($link);
```

Examiners' Guidance Notes

Again, a popular question with over two-thirds of the candidates.

Part a) Well answered.

Part b):

- i) Generally well answered, but many candidates approached the question from the benefits of validation rather than consequences of a lack of validation.
- ii) Generally well answered, but examples given are not always clear as to why and where they should be validated (i.e. at client or server end).

Part c):

- i) Reasonably answered, but a number of candidates were not entirely familiar with the syntax of forming regular expressions.
- ii) Poorly answered in general, with confused and incorrect syntax.

Part d):

- i) Well answered, in a variety of programming languages.
- ii) Variably answered; most candidates were able to perform the database part, but neglected to write a response page after entering the details in the database, or an error page if it failed.
- iii) Weaker candidates did not attempt this part; of those that did attempt it, the answers were generally good, though often neglecting to write a response page for the case in which no videos were found.

Question 3

Explain, with a suitable realistic example in each case, what is meant by:

a)

i) *denial of service* (DOS) attack **(2 marks)**

ii) *cross site scripting* (XSS) attack **(3 marks)**

iii) *buffer overrun* vulnerability **(3 marks)**

b) For each of the three elements (client, network, server) involved in a web transaction over a conventional wired network, detail TWO security risks to sensitive data and, for each risk identified, list the consequences of a breach of security. **(10 marks)**

c) What additional risks do wireless network connections (such as 802.11b) to the WWW bring, over and above those existing on conventional network connections? Outline possible solutions **(8 marks)**

Answer pointers

(a)

i) A denial of service attack causes a loss of services (typically network services) to legitimate users. E.g. Attack on Microsoft.com (2003).

ii) One potential defintion, as referenced from Microsoft.com at http://www.microsoft.com/technet/archive/security/news/crssite.mspx : *"Cross-Site Scripting would potentially enable a malicious user to introduce executable code of his choice into another user's web session. Once the code was running, it could take a wide range of actions, from monitoring the user's web session and forwarding a copy to the malicious user, to changing what's displayed on the user's screen. Even more seriously, the script could make itself persistent, so that the next time the user returned to the web site, the malicious user's script would start running again."*

iii) A buffer overrun (or buffer overflow) vulnerability occurs when a program writes data beyond the allocated end of a buffer in memory, potentially overwriting existing code. This may lead to the program executing arbitrary code. E.g. The Morris Worm (1988), Code Red (2001).

(b) Again, many different risks:

Client-end:

- Hardware Keylogging – Maintain physical security
- Trojaned software - run up-to-date virus checker, or Tripwire
- Forged Emails e.g. as recently pretending to be from Barclays or Microsoft – Common sense

Network:

- Sniffing – Encryption, detection of "sniffers"
- Retransmission – Encryption with timestamping or unique IDs
- Spoofing/Masquerading – Host authentication (perhaps by Public Key)
- Man-in-the-middle – as above

Server-end:

- Server compromise (leading to database disclosure) – authentication, keeping software up-to-date, regular password changing, minimum access privileges, logs, intrusion detection, firewalling, port sentry etc.
- Trojan – Virus checking, port scanning of self

(c) Risks include:

- All conventional attacks with the added disadvantage that there is no need to physically access network cabling
- "Warchalking" to discover wireless hotspots
- Connecting to networks without authorisation
- Monitoring wireless traffic

Possible solutions:

- MAC authentication
- Hiding SSID of access point
- WEP/WPA encryption (but be aware of limitations)
- Wireless "honeypots" to catch abusers
- Physical Security

Examiners' Guidance Notes

Approximately half of the candidates attempted this question and in general the answers were very poor.

Part a):

i) Generally well-answered.

ii) Very few candidates were familiar with the XSS attack, many considering it to be "hacking the scripts on a server."

iii) Typically candidates confused this with a Denial of Service attack, or did not answer it.

Part b): This was poorly answered; Candidates re-stated answers from a similar past exam question, but unlike previous years, this question did not ask for risks and solutions but instead risks and the consequences of a successful attack; a very different aspect.

Part c): Reasonably well answered, but a number of candidates show little familiarity with wireless networking and so were unable to answer this part. Solutions were often quoted verbatim from answer pointers of previous years on WWW:BTB, which does not give a candidate the opportunity to demonstrate their knowledge in anything more than a superficial form.

Question 4

a) List the characteristics of *static* and *dynamic* web pages. **(3 marks)**

i) Describe the technologies and tools used in the creation of a static web page that includes images. **(3 marks)**

ii) Without repeating those elements mentioned in ii) above, describe the additional technologies and tools used in the creation of a dynamic web page. **(4 marks)**

b) `HTMLDocument`, as defined in the Document Object Model (HTML) Level 1, defines a number of attributes.

i) List FOUR key attributes of `HTMLDocument` **(2 marks)**

ii) With reference to `HTMLDocument`, explain how you can replace an image in a web page using Javascript. **(3 marks)**

c) Figure 1 details the source code of a web site for a bookshop.

i) Draw a diagram to complete the missing sections A, B and C indicated in Figure 2 below to illustrate the output of this file when it is first loaded in a browser window. (State the browser you are assuming use of.) **(3 marks)**

ii) The links have JavaScript actions attached to them. Describe what will happen on screen in relation to user interaction with the links. **(4 marks)**

iii) The bookshop wishes to add extra details to the web page, with the same format and functionality as the current content. Write code to enable the menu group displayed in Figure 3 (overleaf) to be generated and displayed. **(3 marks)**

(Note: Your answer should list only the changes and additional lines required)

Answer pointers

a)

i) A static web page has the following characteristics:

- Interaction only through traversing hyperlinks or through plugins.
- Information controlled by the website creator.
- No ability to configure the presentational aspects of the site, except at a local browser level.

A dynamic page tends to have at least some of the following characteristics:

- Can respond to input data provided by the web browser when retrieving the page.
- Can alter or customise the output based on certain factors or criteria.
- Can obtain information from external sources e.g. databases.
- Can be transaction based

ii)

Static development tools and technologies include:

- Web page editor (either a standard text editor e.g. Notepad, or a WYSIWYG editor, e.g. Dreamweaver) for writing HTML.
- Image acquisition and manipulation tools e.g. Scanner, Photoshop
- File transfer software (such as FTP - possibly butlt into the web page editor)
- Web server (e.g. Apache)

iii)

Dynamic development tools and technologies include:

- Scripting languages (e.g. PHP, ASP, JSP, Coldfusion, Perl)
- Possbily development tools/IDE for the dynamic language (e.g. Coldfusion, NetBeans).
- Web server supporting the chosen dynamic languages

b)

i)

DOM:

- title
- referrer

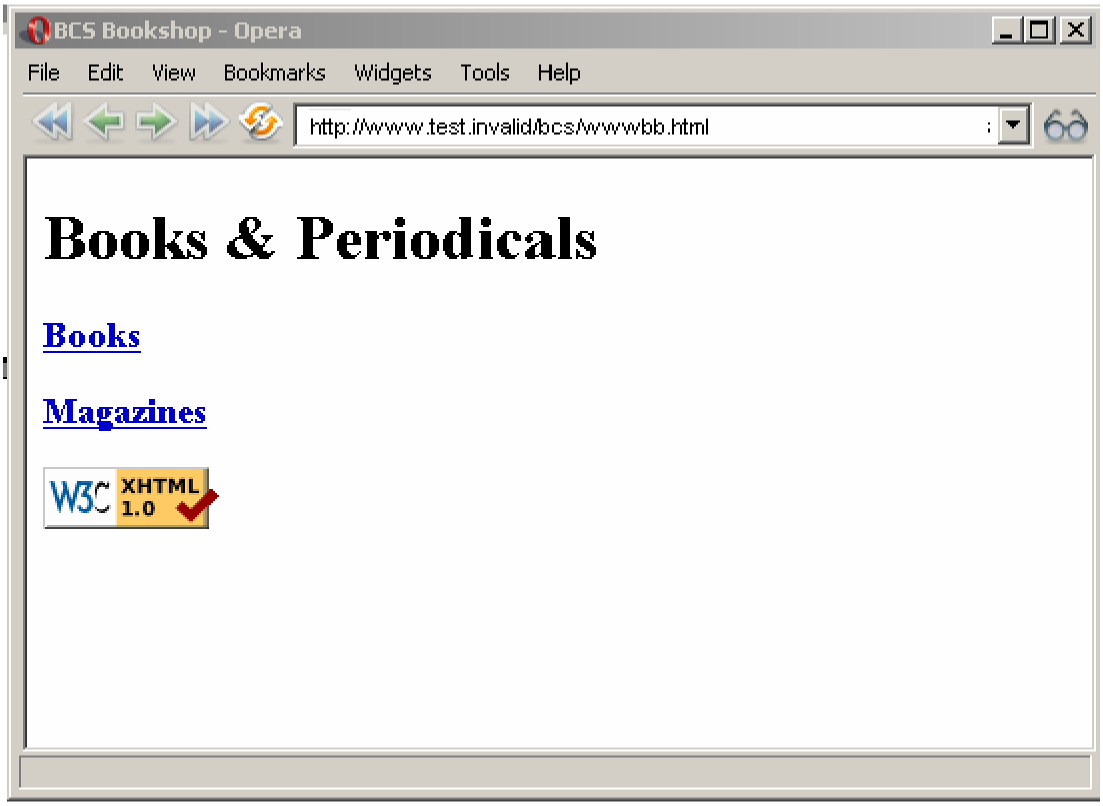
- domain
- URL
- body
- images
- anchors
- links
- forms
- applets
- cookie

ii)

document.images[whatever].src = “http://new_image_url”

c)

i)



Books & Periodicals

Books

Magazines

ii)

When the user clicks a link, the related menu group is displayed.
If the mouse is clicked on the link a second time, the menu is made invisible.

The menu groups are displayed and made invisible independently of each other.

iii)

Add the following code after line 47:

```
<h3><a href="page3.html" onmouseover="return togglegroup('group3')">
    Journals</a></h3>
<p id="group3" class="group">
    IE Proceedings<br />
    MCQ<br />
    Clinical Review
</p>
```

Examiners' Guidance Notes

Just over half of the candidates answered this question.

Part a): Most candidates answered this appropriately, however the good candidates clearly demonstrated the technologies and the respective tools within those technologies. There were a number of candidates who failed to structure their answers.

Part b):

i) A number of candidates failed to understand that the question was about DOM and interpreted it as though it was about basic HTML which resulted in answers discussing and demonstrating HTML tags.

ii) A number of candidates went onto write JavaScript functions to enable image swap when in fact it was simply asking about JavaScript action with a DOM reference.

Part c): As with question 3, quoting from past exam papers was the downfall of many candidates here; the two menu items operated independently and the action was triggered with a mouse click as opposed to being linked and triggered on a mouse over.

Question 5

a)

i) With a specific example, explain what is meant by the term RSS. **(2 marks)**

ii) What is meant by the term *podcasting*? **(2 marks)**

iii) With reference to real life examples, explain what benefits podcasting and RSS-supported sites offer over traditional broadcasting media. **(5 marks)**

b)

i) Define and explain the terms *VoIP* and *IM*. **(3 marks)**

ii) With specific reference to real-life and contemporary VoIP and IM applications, explain how Internet-based real time communications have changed the way people communicate, both at work and socially. **(5 marks)**

c) Explain, with specific examples, how the WWW has transformed the way that people shop for:

iv) Music

v) Air travel

vi) Financial Services **(10 marks)**

Answer pointers

a)

i) Really Simple Syndication (or equally, Rich Site Summary or RDF Site Summary) - a family of XML file formats for Web syndication. E.g. RSS feed on BBC News to give up-to-date summaries.

ii) Podcasting is a method of distributing multimedia files (generally audio files), mainly geared for some form of personal audio player (PC, mp3 player, etc.) over the Internet using syndication mechanisms such as RSS. An example would be the use of Podcasting by Professor John Fothergill at the University of Leicester, the New Scientist podcast, etc.

iii) Any sensible reason will be considered that compares podcasting with traditional media (radio, TV, magazines, papers, etc), but as an example consider e.g. as from http://spinfluencer.blogspot.com/2006/04/benefits-of-podcasting.html

- *“Allows listeners to time-shift and place-shift media consumption*
- *100% efficiency, since episodes are only downloaded by listeners on an opt-in basis*
- *Easily accessible to a global audience that is not defined by geographic boundaries*
- *Access to an educated, influential audience with a high disposable income*
- *Ability to leverage electronic programming without an outside news media filter*
- *Most cost effective electronic media distribution channel available”*

b)

i) Any sensible reason will be considered that compares podcasting withVoIP (Voice over Internet Protocol) is the routing of voice (rather than “data”) over an IP-based internet such as the Internet.

ii) IM (Instant messaging) is the name given to a suite of applications based on IRC ideas that allow fast point-to-point messaging, usually between 2 people, but with the possibility for group communication.The discussion should make reference to the benefits of VoIP and other instant communication methods over and beyond emails and other communication methods e.g.

- Real time group communications/remote meetings.
- Location of offices becomes less important (e.g. communicating Internationally).
- “Free” calls using VoIP from computer to computer.

- Videoconferencing using web cams.
- Reference to related issues such as the use of real time communications over PDAs (e.g. Contact 3.0 for explorers) will be considered.

Reference to real life IM providers (e.g. ICQ, AIM, MSN) and VoIP providers (e.g. Skype) are required for full marks.

c)

i) Aside from the general issues of global market etc, the discussion should make reference to the advent of mp3 and file distribution networks, commercial ventures such as Napster and iTunes, ordering CDs from abroad (e.g. from Play 247), and marketing and releasing of material direct from artist (e.g. mp3.com).

ii) The discussion here should look at the concept of e-ticketing, booking actual seats (instead of a general reservation), checking in and confirming online, and broker sites like expedia.co.uk.

iii) The discussion here should mention subscription-based web access e.g. credit reports, access to journal articles, support contracts over the web, topping-up of mobile phone credit, transport (e.g. oyster) with online purchase and renewal.

Examiners' Guidance Notes

Two-thirds of candidates provided answers to this question.

Part a):

i) Generally well answered.

ii) Generally well answered.

iii) Those who understood the idea of podcasting were able to answer this question exceptionally well, whilst those that did not typically did not attempt this part.

Part b):

i) Generally well answered.

ii) Reasonably answered from a work perspective, but the coverage of social aspects were much weaker.

Part c):

i) Generally well answered.

ii) Candidates did not show much familiarity with using WWW to purchase air travel, except for checking timetables online.

iii) Poorly answered. Students discussed the use of the WWW to check bank balances or transfer money; this is not relevant to the question of shopping for financial services.