

ACM Pacific NW Region Programming Contest
10 November 2001

PROBLEM G

WHAT'S MY SIZE?

When designing an HTML page for display on the World Wide Web, the length of time it takes to display the page in the user's browser is a major consideration. Studies show that most users will not wait more than 10 seconds for a page to download before "clicking out", and costing the web site valuable "eyeballs".

Many current HTML editors have the ability to let the designer know page download time for the "average" user. Your task is to write such a utility to compute download time, assuming a "static" HTML document. (For our purposes, a static HTML document is one that contains no multimedia, Java applets or dynamically generated content. It consists of only HTML code and images.) Download time will be determined by the total size (in bytes) of all images to be displayed in the document, plus the document size (in bytes, 1 character = 1 byte), which is computed by its character count. This total number of bytes is used to determine the download time, assuming 1000 bytes = 0.2 second.

For example, given a simple HTML document such as this:

```
<HTML>
<HEAD>
<TITLE>ACM</TITLE>
</HEAD>
<BODY>
<IMG SRC="acm.jpg"><BR>
<IMG SRC="contest.jpg">
</BODY>
</HTML>
```

and assuming the size of "acm.jpg" is 10k bytes and "contest.jpg" is 5k bytes we can compute the download time as follows:

- Number of bytes in document: 112 (assume there is a CR/LF character at the end of EACH line)
- Number of bytes in images: 15000
- TOTAL SIZE: 15112 bytes
- DOWNLOAD TIME: 3.02 seconds

For this problem, we are going to add an attribute to the IMG tag, called "size". This attribute will have an integer value, denoting the image file size in bytes. For example, the sample above would look like this:

```
<HTML>
<HEAD>
<TITLE>ACM</TITLE>
</HEAD>
<BODY>
<IMG SRC="acm.jpg" size="10000"><BR>
<IMG SRC="contest.jpg" size="5000">
</BODY>
</HTML>
```

ACM Pacific NW Region Programming Contest

10 November 2001

Note that if an image is used more than once on the page, it is only counted once in determining download time, since it won't have to be downloaded again, once it's been downloaded originally. (Images with the same file name, but in different directories are NOT the same image. Image names ARE the same if a case-insensitive string compare declares them to be equal.)

Note that HTML tags are NOT case sensitive.

Note that other HTML tags besides the tag can contain the "SRC=" attribute and the "size=" attribute.

Note also that an tag can contain the "SRC=" attribute along with other attributes, and these can be in any order. For instance, for our purposes, all of these are valid tags:

```
<IMG SRC="acm.jpg" size="10000">  
<img BORDER="0" SRC="/images/contest.gif" size="5000">  
<Img HEIGHT="42" WIDTH="42" size="5000"  
src="../Images/Answers/Answer.jpg">
```

(You may assume that the path/image name and size value are surrounded by double quotes.)

Specifically, you may assume that the only spaces in a tag will be either between name=value pairs or embedded in quoted strings. More specifically, there will be no spaces between the < and tag name, around the equal signs, before the >, or anywhere else not already specified.

Input:

Input for this problem will begin with a single integer on a line, denoting the number of HTML files to process. Following this line will be a series of document listings. Each HTML document listing will begin with <HTML> tag and end with </HTML> tag (not case-sensitive!), and will contain 0 or more image tags. There will be no blank lines within the HTML document. The ending </HTML> tag for each test document listing will be the last item on a line. Image file names (including the path) will not exceed 80 characters.

Input for this program will be the text file **G.in**

Output:

Output consists of a series of lines, one per HTML document, listing the time, in seconds, to download that document, rounded to 2 decimal places.

ACM Pacific NW Region Programming Contest

10 November 2001

Sample I/O:

Input:

```
2
<HTML>
<HEAD>
<TITLE>ACM Contest</TITLE>
</HEAD>
<BODY>
<HR><H3>IBM Sponsors ACM-ICPC!</h3>
<IMG SRC="acm.jpg" size="5000"><BR>
<p>A new paragraph</p>
<center><img BORDER="0" size="1000" SRC="/images/contest.gif"></center>
&nbsp;
</BODY>
</html>
<HTML>
<HEAD>
<TITLE>Pacific NW Region</TITLE>
<script language="JavaScript" src="../../scripts/utils.js"></script>
</HEAD>
<BODY>
<b>PacNW Rocks!</b><br>
<hr>
<table>
<tr>
<Td><IMG size="5000" SRC="../../IBM.jpg"></td>
<td><p>A new paragraph</p></td>
<td align=center><img BORDER="0" SRC="/images/pacNW.jpg" size="7000">
</td></tr></table>
<br><br>
<br>
<h1>Bye!</h1></BODY></HTML>
```

Output:

```
11.26 seconds
7.50 seconds
```