

Argument

Separating Content, Presentation and Behavior Components in the Census GIDS Project

Prepared by Steve Schafer, Ph.D., CTO
Fenestra Technologies Corporation

Content is important to both computers and humans, whereas layout is important only to humans. Interspersing layout with content makes it more difficult to work with only the content or only the layout.

For example, you can go to the UPS web site and track the status of a package; all you have to do is enter your tracking number. But let's say you want to automate this by writing a program that goes to the web site for you, and then stores the retrieved status information into a local database. As it is right now, the content of the web page (the status information) is all mixed up with the layout information (the stuff that makes the content palatable to human viewers). If UPS decides to change the layout, even slightly (to improve the appearance of the page, for example), it's likely that your program will stop working, because it won't be able to understand the new way in which the content is interspersed with the layout.

By sending the content separately from the layout, things that are interested in only the content, like your program, can concentrate on that and ignore the rest, and have a much better chance of being able to continue working even if the layout changes.

There's another thing to be gained by separating content from layout: it becomes possible to label and identify the various bits and pieces of the content, so that a computer program can determine what is what.

Let's visit the US Postal Service web site and retrieve the standardized address for Fenestra. A portion of the web page returned by the server looks like this:

```
<b>
20410 CENTURY BLVD STE (EVEN
Range 100 - 230)<BR>
GERMANTOWN MD 20874-1187<BR>
</b><TABLE
border=0><TR><TD><i>Carrier
Route</TD><TD></i><b> : C053
</b>
</TD><TD><i>
County</TD><TD></i><b> : MONT-
```

```
GOMERY          </b>
</TD></TR>
<TR><TD><i>Delivery
Point</TD><TD></i><b> : 99</b>
</TD><TD><i> Check
Digit</TD><TD></i><b> : 4</b>
```

If you look closely, you can find the necessary information, but it's embedded in the page as a whole, without being identified, and the various fields are concatenated in a way that makes sense to a person, but not a computer.

Now see how it might look if encoded in machine-readable XML:

```
<standardized-address>
<address line="1">20410 CENTURY
BLVD STE (EVEN Range 100 -
230)</address>
<city>GERMANTOWN</city>
<state>MD</state>
<zip>20874-1187</zip>
<carrier-route>C053</carrier-route>
<county>MONTGOMERY</county>
<delivery-point>99</delivery-point>
<check-digit>4</check-digit>
</standardized-address>
```

As you might imagine, this format is much easier for a computer to digest since everything is in a simple, content-oriented format, and all of the information is labeled as to its meaning.

Another scenario in which separating content from presentation is useful is a web browser that is adapted to the needs of people who are visually impaired. If the content of a web page were separated out and identified as to its significance, the browser could make intelligent decisions about how to present that content to the user.

In the Census GIDS project, the separation of content and layout is important because the content is used in more than one place (i.e., in both paper forms and electronic forms).

Keeping the content separate from the layout makes it simpler to reuse the content in the two different contexts.

Shaping
Dynamic
Content

