Common Errors

This section details common errors in HTML composition, that may lead to documents which are not fully device-independent. The behaviors of these errors are undefined, so certain browsers may render them as intended **but not all browsers are guaranteed of doing so**. Therefore, these mistakes should be avoided, even if your browser of choice renders your documents correctly.

Contents

Paragraph Break Errors

This is probably the most prevalent kind of error, and is the number one culprit in cases of ugly HTML rendering. If you fix nothing else, fix these! Perhaps the biggest misconception about the <P> element is that it signals an end-of-paragraph, rather then a paragraph break. According to the specification, "<P> is used between two pieces of text which otherwise would be flowed together".

In most cases this is not important -- functionally, the <P> serves as an end-of-paragraph marker. However, in certain contexts, use of <P> should be avoided, such as directly before or after any other element which already implies a paragraph break. To wit, the <P> element *should not be placed either* **before** *or* **after** *the headings, HR (can I get a ruling on this? people don't handle HR consistently... X Mosaic has no white space before or after, and Lynx appears to put white space after), ADDRESS, BLOCKQUOTE, or PRE.*

It should also not be placed immediately before or after a list element of any stripe. That is, a <P> should not be used to mark the end-of-text for <LI>, <DT> or <DD>. These elements already imply paragraph breaks.

Caveats

Some clarifications on the above might be in order. One is the the difficulties of rendering appropriate white space by a browser. While it is true that all of the entities mentioned above imply a paragraph break, this only occassionally means that they also imply white space between sections -- this depends on the browser. So, while you might feel inclined to add a <P> in order to fix white space problems, please think twice and avoid it if you can.

Also, when using the glossary list (DL), please try to avoid using multiple DD's (definitions of terms) in order to provide multiple entries for a term (DT). Instead, use a <P> marker between paragraphs in a definition. The use of a DD (definition) without a matching DT (term) is illegal, although a DT without a DD can be used without dire consequences.

All clear now?

Character and Entity Reference Errors

Simply put, a character reference and an entity reference are ways to represent information that might otherwise be interpreted as a markup tag. For instance, in order to represent <P> in this text, I had to use `<P>` in my raw HTML. There are currently five entities for this purpose in HTML, as well as several entities which allow encoding of the ISO Latin-1 Character Set.

The most common error in the use of references is to leave off the trailing semicolon. Also, no additional spaces are needed before or after the entity/character reference.

URL Errors

Another misunderstood aspect of HTML is in the composition of URL's.

Directory Reference Errors

One grey area involves references to directories. It is possible to request an index of a directory from an HTTP server. The typical response from the server is to either return a pregenerated index document (which is often the document "index.html" in the referenced directory), or to construct an HTML document on the fly which contains a listing of all files in the directory. However, when making such a directory reference, *it is important to make sure to have a trailing slash on the URL*. That is, if you were to request the index of the directory which this document resides in, you would want to refer to it as http://www.willamette.edu/html−composition/, not as http://www.willamette.edu/html−composition.

Some servers are able to catch these errors, and provide redirection to the proper URL, but it's best to get the URL right in the first place −− notably because not all browsers support transparent redirection.

Not Using Fully Qualified Domain Names

Problems can arise when the hostnames in URLs aren't fully qualified In local networks, you can usually refer to your own machines simply by their names −− for instance, here at Willamette we refer to our local WWW server as "www". However, the server's FQDN (fully qualified domain name) is "www.willamette.edu". The FQDN provides enough information that any host, anywhere on the Internet, can find this particular machine. (It's like trying to find all the Vermeers in New York :).

What happens is that an HTML might construct a link that looks like this:

```
<A HREF="http://www/~jtilton/metanoia/">Metanoia -- A Change In
Spirit</A>
```

which produces a link to Metanoia −− A Change In Spirit that will only work for people in the local network that that machine is on. A correct link would look like this, instead:

```
<A HREF="http://www.willamette.edu/~jtilton/metanoia/">Metanoia</A>
```

which would allow all of you who are interested in Metanoia to actually follow the link.

This leads almost directly into:

Improper Use of Relative URLs

Finally, a brief section on relative URLs. It is possible to construct a "relative" URL, which gives you the following advantages:

- It's shorter.
- It makes a collection of documents which are linked together more portable (easier to move from directory to directory, or server to server).

However, relative URLs can also break things.

A relative URL is a URL which doesn't contain all the necessary parts of a "full" URL (scheme, host, path information). There's a large number of things which might fit this description! The browser will try to assume the parts that have been "left out" by using the information from the URL of the document which contains the link. However, not all browsers will make these assumptions in the same way. Here's a short list of what's "safe" and "unsafe" (based on experience, and not on a specification anywhere −− unfortunately).

Safe: Same directory relative URL's

A reference to a document in the same logical directory (such as <A HREF="strict−html−gp.html">Good Practices</A>) is safe. This kind of reference, roughly speaking, contains no "/"'s.

Safe: Same server relative URL's

A reference to a document in the same server (such as <A HREF="/~jtilton/">Eric's Hyplan</A>) is also safe. This kind of reference, roughly speaking, will begin with a "/". (It will also be semi−absolute, in that it starts at the top of that server's directory structure...)

Unsafe: Most other kinds of relative URLs

References such as <A HREF="~jtilton/euphonium.html"></A> can be dangerous −− some servers will interpret that as meaning "go up one directory level, find the directory '~jtilton', and then find 'euphonium.html' in it." And some won't. Roughly speaking, these are references that contain a slash somewhere in the middle, that aren't "full, absolute" URLs.

Unsafe: "file://localhost/..."

It's also possible to have a reference to "file://localhost/some/file/pathname". What this does is references the file described on the *local host of whoever is browsing the document.* Which is why a reference to <A HREF="file://localhost/etc/motd"></A> will display the message of the day on your machine, not the message of the day on my machine. Unless you know what you are doing, these references will really mess up your documents.

(This sub−section isn't written very well, I fear. If anyoe has any better copy, I'll gladly put it here instead. −et 4/7/94)

Missing Quotes in Start Tags

One common error that I used to make all the time (I use Marc Andreesen's html−mode.el for Emacs these days −− I had to learn Emacs, but now it's so much easier to write HTML!) was to leave off a quote in my start tags. For example, this reference to the euphonium, king of instruments should look like:

```
<A HREF="http://www.willamette.edu/~jtilton/euphonium.html">
```

but I would often use

```
<A HREF="http://www.willamette.edu/~jtilton/euphonium.html>
```

instead. I suppose by the end of that huge URL, I'd forgotten it was supposed to be quoted. The behaviour of browsers upon encountering this varies −− some display a proper link, but you can't follow it, while others actually eat up huge portions of the following text, thinking it to be part of the URL.

Missed End Tags

Many of the HTML elements contain information within them. For example, emphasized text would be rendered as *emphasized text*. There is a start tag (<EM>), some content (which may include text, and in some cases, other nested elements), and an end tag (</EM>, indicated by the </). A common mistake is to miss the / in the end tag. All elements (except empty elements, see next paragraph) must be terminated by an end tag −− otherwise, undefined behavior may occur.

Some HTML elements may be empty, such as <P> and <HR> (CERN provides an extended discussion on element content). If this is the case, there is no need for an end tag.

You may also want to

- Go on to "Things to Avoid"
- Go back to "Good Practices"
- Return to the table of contents

Last modified: Apr 7, 1994

James "Eric" Tilton, HTML Guru Wannabee, jtilton@willamette.edu