

Appropriate use of content formatting elements in a content managed portal environment

By Thomas M. Farrell

Senior Web Technology Consultant

tom@tomfarrell.org

Copyright 2005 Thomas M. Farrell, all rights reserved. Unlimited license is granted to employees of the Commonwealth of Massachusetts to reproduce this document in electronic or printed media for internal use.

Introduction

Content management has become a standard tool in the contemporary web industry. However, while many sites are implementing it, few of the impacted content creators have experience with its unique limitations and benefits, and how to leverage their use of content formatting options to improve the quality and usefulness of their results.

Many common web practices are non-optimal and can result in degraded-quality experience for blind users, sub-optimal search engine placement, and a generally unpleasant experience for the average user.

This document examines specific practices of this problematic type to demonstrate how they can be problematic and what alternative practices exist that can easily resolve the problem through more appropriate use of standard HTML content formatting elements.

Audience

The intended audience of this document are persons who have at least incidental experience creating or managing content in a web context and have some professional responsibilities relating to web content in a content managed or portal environment, but who may not be full time web professionals.

Formatting different parts of one word separately

The problem

Sometimes it's tempting to format parts of one word differently from the rest of the word. For example, the first letter could be made larger and bold, as shown:

There

This is bad because, in the HTML, it breaks up the word with formatting instructions. In this example, the HTML which forms the single word "hello" would look something like this:

```
<font size=+2><b>T</b></font>here
```

What this does to the blind

Some (not all) screen readers will interpret the "T" as being a separate word from "here" because in the HTML they're separated by formatting instructions, and pronounce it as "Tea here" instead of "there".

What this does to search engines

Some search engines will interpret the "T" as being a separate word from "here" because in the HTML they're separated by formatting instructions. It would be indexed as "T" and "here" separately, and therefore the page would be less strongly associated with the word "there". This would cause the page to rank lower in searches for "there" than it actually should. This makes the page harder for users to find.

What this does to average users

Average users will find this to be suboptimal because their eye has to adjust to formatting changes within the individual word. Extended reading of variable type within individual words will be more tiring for the user than plain text.

Also, if you use extensive markup to stylize the appearance of text, such as making the first letter of every paragraph bold, it makes your

content look different from everything else on the web site. This confuses users, who generally like everything to be consistent. By eliminating extraneous formatting, your pages will look more consistently similar to other pages on the site, which will make the user's experience more comfortable.

The correct way to do it

The correct way to use formatting to handle this situation is, never format part of a word differently from the rest of the word. This will prevent the confusion and difficulties with screen readers and search engines, and will make the text easier to read.

Use of “click here”

The problem

The phrase “[click here](#)” is commonly used as the text of hyperlinks to indicate to the user where to click.

What this does to the blind

Many blind users have a screen reader application for their web browser software which reads all the hyperlinks from the page to them at the beginning of the page, the end of the page, and whenever they press a specific key. This makes a “verbal menu” of the links for them, so they can locate the link they want and use it.

If all the links on your page contain the text “click here”, what they'll hear when they try to use your links is “click here click here click here click here click here click here...” and that makes the page unusable for them. Even if you have several links and only one of them says “click here”, it gives the link no context, so they have to memorize the location of the link to figure out which one it is they want to follow, instead of just figuring out which one they want from descriptive text.

What this does to search engines

In determining how closely a page should be associated with keywords, search engines such as Google not only examine the text of the page itself, but also the text of the links which point to the page. Words that other pages use about a page can, in fact, be considered by the search engine to be more important than the text of the page

itself. As a result of this, a notorious incident occurred in which entering the phrase “more evil than satan himself” into the Google search engine would return, as its top result, a link to Microsoft.

So, if you use “click here” to link to a page, the page you linked to becomes associated more closely with the phrase “click here” than it should be, and less associated with more appropriate keywords than it should be. This causes the target page to be incorrectly ranked in search results, which could make it harder for users to find.

What this does to average users

Savvy web users frequently see the phrase “click here” used on poorly designed, amateur web sites. They therefore have a certain tendency to discount the value of the information your pages provide if you use the phrase “click here”.

Some elderly users with little computer experience become confused when physical action verbs are used as a metaphor for actions in the user interface. So, when they see the phrase “click here” they attempt to press the screen there with their finger, which of course won’t get them the desired result.

The correct way to do it

Instead of “click here”, place your hyperlink on contextual text. For example:

[Click here](#) to view a list of available documents.

...could be rephrased as:

View a [list of available documents](#).

...or as a bulleted item:

- [Available documents](#)

...or could be displayed as a button in a form. If you absolutely have to even refer to or describe clicking something, try to use the word “select” instead of “click”. Users view it as more sophisticated.

Use of URLs as hyperlinks or in visible text

The problem

In creating web content, it is frequently tempting to make the text of a hyperlink the same as the URL of the hyperlink, so that viewers will be able to see the URL in case they want to write it down or print the page and type the URL in later.

What this does to the blind

Blind users utilize screen reader software which reads the contents of the page to them. URLs are very slow for the software to pronounce because the software determines it's unpronounceable and reads only parts of it. So, a URL such as:

<http://www.tomfarrell.org>

may look simple enough to an ordinary sighted person, but the screen reader may pronounce it as something like:

aitch tea tea pee colon slash slash double you double you
double you period tomfarrell period org

This causes the page to take longer to read. If several such instances appear on the site, then it's even worse. And, if the text appears in a hyperlink and the software is set to read all the links before and after the page, they have to sit through that jumble of verbal code three times.

What this does to search engines

As described above, search engines consider the text used to link to a page to be very important. If you link to a page with its own URL as the link text instead of contextually appropriate text, it fails to gain association with appropriate keywords in the eyes of the search engine. This could cause it to be ranked lower in search results, which would make it harder for users to find it.

What this does to average users

Average users don't like to have to think about URLs, they'd rather just see your content. Some don't even know how URLs work, they just use the search bar in their browser whenever they want to get to

something and click through to pages in the search results, and bookmark what they like. So, using URLs could confuse them.

The correct way to do it

Always use descriptive text as the text of a link instead of the URL. If the user really wants to keep the URL, they can bookmark it. If they're not at home, they can select the link and write down the URL from the browser window while viewing the resulting page. So, there's no real need to put a URL in the readable text of the page.

Leaving URLs out of the readable text also makes for a cleaner-looking page, because the user only has to read your valuable content, not computer codes.

Use of bold, italic, and/or underlined text instead of Header 1, Header 2, and Header 3 elements

The problem

In creating web content, documents are often created in a word processor, then exported to HTML, which may be cleaned up a little and used or may be used as-is.

Section headers in the original document may have been created through the use of some combination of different font sizes, bold, italics, and underlines. When the document is translated to HTML, these formatting methods are preserved, instead of converting the formatting to a Header 1, Header 2, or Header 3 (H1, H2, or H3) element.

What this does to the blind

Some blind users use screen reader software for their web browser which can summarize a page by converting the header elements to a kind of "verbal outline", which allows them to select only the most interesting part of the page to hear. Without this feature, long pages can take quite a bit of time for them to examine, because they have to listen to the computer read it all to them, they can't "skim" the page as a sighted person would.

If the page's section headers are not formatted as header elements, but are instead formatted manually using visual markup elements, the

browser can not summarize the page, forcing blind users to listen to the whole thing to find what they want.

What this does to search engines

Search engines index pages by creating a list of words with which to associate the page, and providing a “weight” for each word to determine how strongly the page is associated with that word. More appearances of the word help weight the page more strongly toward that word, which makes the page rank more highly in search results for users searching for that word.

Words which appear in header elements (H1, H2, H3 etc) are typically weighted more strongly for the page, because the search engine assumes that words in headers will have stronger bearing on the content of the page than random average words on the page.

If there are no header elements on the page, no words are weighted more strongly. This represents a lost opportunity to increase the page’s rank in search results for the words of the headers.

What this does to average users

Users tend to quickly adjust to the overall look and feel of a web site. And, most sites do not make enormous adjustments to the appearance of standard header elements, so users become accustomed to what they look like across differing sites.

When your pages don’t use standard header elements for their section headers, users will sometimes fail to recognize the page’s “headers” as such, and will fail to understand the structure of the page correctly.

A related maintenance issue

Many web sites now use stylesheets to adjust the look and feel of the site as a whole. If your pages don’t use standard header elements, their headers will not receive ongoing adjustments which may be made by the site management.

The correct way to do it

When converting content from other formats, make sure to convert all page and section headers into relevant header elements.

Using a series of underscores or hyphens to make a line

The problem

In a typewriter, to make a line, you would simply type a series of hyphens or underscores. Many people do the same thing when creating pages in a word processor or for the web.

What this does to the blind

Some screen reader software will attempt to *pronounce* a series of hyphens or underscores. Usually it's pronounced as an extended pause. This can be very confusing for a blind user when they're trying to listen to the content of a page and there's a long pause for no apparent reason.

Related maintenance issues

Many web sites now use stylesheets to adjust the look and feel of the site as a whole. If your pages use hyphens or underscores to make lines instead of a standard Horizontal Rule (HR) element, these lines will not be correctly formatted according to the site's stylesheet standards.

Also, a series of hyphens or underscores may cause awkward aesthetic problems. For example, it may be too short to fill the desired area, or it may be too long and create an undesirable "hanging" piece of extra line, like this:

--

The correct way to do it

Using a standard horizontal rule element will always perfectly fill the specified space, will never create "hanging pieces", and will obey relevant stylesheets.

Use of unnecessary graphics

The problem

It's frequently tempting to put graphics on a web page to jazz up its appearance. If there are no available graphics which add value to the content of the page, purely decorative graphics may be used.

What this does to the blind

In order to conform to World Wide Web Consortium (W3C) practices for web site accessibility for the blind and visually impaired, every graphic's HTML tag must have an "alt" parameter. This is an element which provides descriptive text which is substituted in place of the graphic when the page is converted to pure text for persons using a browser which does not display graphics, or using a screen reader.

For example, if you have a photograph of a teddy bear, its "alt" parameter should read "Photograph of teddy bear."

If you have strictly decorative graphics, and if you make your page conform to correct practices for accessibility, you'll be wasting your blind users' time with descriptions of pretty pictures, instead of getting them promptly to your content.

What this does to search engines

Descriptive "alt" parameters for purely decorative graphics cause the page to become more closely associated with the potentially irrelevant words of the descriptive text for the images rather than with contextually appropriate keywords. This could negatively impact the page's rank in search engine results, making it harder for users to find. In other words, the descriptive text of the useless images *dilutes* the text of the real content.

The correct way to do it

When considering an image for your page, consider if it actually contains information or is purely decorative. If it's purely decorative, it's best left out. Leave it up to the site's management to make the pages look pretty using attractive templates.

If the image contains information, consider whether you can as easily convey the same information using text. If you can, why not just use text? Text loads faster anyway, and better supports the blind.

Unnecessarily active graphics

The problem

There are many cute graphic elements available on the web. For example, many sites use an icon of a letter which folds itself up and stuffs itself into an envelope that files into a mailbox, to indicate where to click to send email.

What this does to the blind

If you use a very active graphic, accurately describing it in “alt” text for the blind can be excessively long. If you make the description too short, you’re not conforming to accessibility standards. Do you really want to distract your blind users and waste their time listening to a description of the cute animated picture?

What this does to average users

Hyperactive graphics continually draw the eye, making it hard for the user to focus on the important content of your page and keep their eye away from the flashy graphics.

Aesthetic considerations

Most professional web graphic artists are appalled by flashy, hyperactive graphics. They sometimes refer to such images as “the spinning flaming logo syndrome”.

Users frequently see hyperactive graphics in two contexts: advertising, and in poorly designed, amateur sites. They therefore consider unnecessarily active graphics to be a sign of lack of quality.

Legal issues

Many graphics of this type are found on web pages of “free” graphics and recycled for use on other pages. Often, these pages of “free” graphics were created by someone who doesn’t understand copyright law and believes that if they can copy it, it’s legal, so they steal other people’s graphics work and tell you it’s “free”. Using such graphics can create liability.

The correct way to do it

Don't use an animated image when a non-animated image will do.
Don't use an image when plain text will do.

Fully justified paragraphs / specifying paragraph alignment

The problem

Many people like to specify that paragraphs should use fully justified alignment instead of left justified alignment, because it creates aesthetically pleasing paragraphs.

What this does to the average user

Studies of reading behavior and eye strain indicate that fully justified paragraphs cause more eye strain than left justified paragraphs, because the eye has to adjust to continually changing spacing.

Also, by specifying the alignment of your paragraphs instead of leaving it to stylesheets, you could cause your pages to appear different from all the other pages on the site. This may confuse some users.

A related maintenance issue

Specifying the desired alignment of your paragraphs in your HTML code could override stylesheets, making it impossible for the site's management to style page content correctly with stylesheets.

The correct way to do it

Just don't specify how you want your paragraphs aligned. Your site's management can handle it for you with stylesheets.

Fixed widths on tables

When creating content for web pages containing information laid out in a table, we frequently specify how many pixels wide the table should be in order to give it a pleasing appearance.

In a content managed environment, we may no longer control the template of the page on which our content appears. The area in which the table appears may be adjusted to be a different size, which

may make the table either too large to fit in the new area, or so small that it looks out of place in a newly enlarged area.

It is therefore best, when possible, to either specify the width of your tables using percentages, or to simply not specify the width so that the table can take up as much of the available space as it needs. Using these methods, the table should resize itself to fit.

Summary

In a content managed environment, less is more. Overly formatting your content tends to create problems, while simplistic use of standard formatting elements tends to create benefits. It is therefore in the best interests of everyone – the management of your site, your users, and you – to simplify your content formatting.