Carnival of HTML



Copyright © 2010 Jerry Stratton

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.1. A copy of the license is included in the section entitled "GNU Free Documentation License"

Introduction	1
What is HTML?	
The web site	
What is that cover?]
The basic web page	···· 3
HTML and BODY	
Paragraphs and HeadlinesThe HEAD of the document	4
Title	
Description	
Keywords	6
Style sheet	6
Paragraphs and Blocks	7
Headlines	7
Finish the paragraphs	
Lists	
Character styles	/ 11
Emphasis	
Links	
Validation	
Document Type	
Character set	. 13
Validation	. 13
Flair	15
Style Sheets	15
Classes and IDs	
Styles	
DIVs and SPANs	
Images	ر1. 16
Ålternate text	
Pullout	. 17
Tables	19
Lists	21
More about URLs	22
Deleted and inserted text	23
Entities (special characters)	·· 24
Ellipses and m-dasnes	. 24
Quotes	
Other characters	. 25
Ampersands	
Extra Credit	27
Special tags	27
Meaningless markup	27
Why XHTML 1.0?	27
HTML 4.0.1	28
Blog comments	28
More Information	
Links	
GNU Free Documentation License	70

Introduction

What is HTML?

HTML is HyperText Markup Language. Hypertext because it is more than text, and markup language because it is a language for marking pieces of text. HTML is the language of web browsers. Using HTML, you describe how your document is structured so that web browsers can display it appropriately. Unlike normal desktop publishing, with HTML you only work in generalities, if you know what you're doing. Rather than specifying exactly what your document looks like, you specify which parts of the document are important, and in what way they're important. The reader's browser then takes that information and creates a web page, regardless of whether that browser is a graphical browser on Windows, a text-based browser on Unix, or a voice reader for the blind.

When writing HTML, you surround various parts of the text with descriptions of what added meaning you want the text to convey. For example, if you want a word to be emphasized, you surround that word with the 'emphasis' HTML code. Almost all HTML 'markup' is done by surrounding the words with the code that affects it. The beginning tag is always a word, such as "em", surrounded by the greater than and less than symbol: . The ending tag is the same thing, but with a slash added: .

There are two forms of HTML: HTML and XHTML. I'll be using XHTML here, but will try to point out the differences with HTML, and why you would use one or the other.

The web site

You can find the latest version of this tutorial, as well as the resources archive, at http://www.hoboes.com/NetLife/Web_Writing/.

What is that cover?

It's from the 1900 Mardi Gras. It has nothing to do with HTML; in fact, it is everything that HTML should not be: crowded, gaudy, and incomprehensible.

The basic web page

Copy the file "Carnival.txt" in the Resources folder to "Carnival.html" and put it in the Workshop folder along with all of the images (we'll get to those later). Open Carnival.html in a text editor, such as Smultron on Mac OS X. It's a reasonably formatted text file, but it certainly isn't what you'd expect to view on the web nowadays.

Go ahead and view it in your browser. It should look like a mass of text, with no text standing out from any other text.



Over the course of this tutorial we're going to make parts of the text stand out: headlines, paragraphs, links, emphasis, and lists. We'll do this by telling the browser what each of these bits of text mean, structurally.

HTML and BODY

Almost everything in HTML is a tag describing the meaning of text. Even the web page itself needs to be surrounded with a tag saying that this is a web page. That tag is the HTML tag. At the very top of the document, type "<html>". At the very bottom of the document, type "</html>".

This is what HTML tags look like: a tag name between angle brackets surrounding some text, and then the same tag name with a slash in front of it to end the text. Use all lower case for your HTML tags. HTML recommends it, and XHTML requires it.

The main part of your web page—the part that people actually see when they're visiting at your web page—is the *body* of the document. Surround all of the text—inside the HTML tags—with "
body>" and "</body>". The body is where the meat of the document goes. All of the information that you're giving to the reader goes in the body.

Paragraphs and Headlines

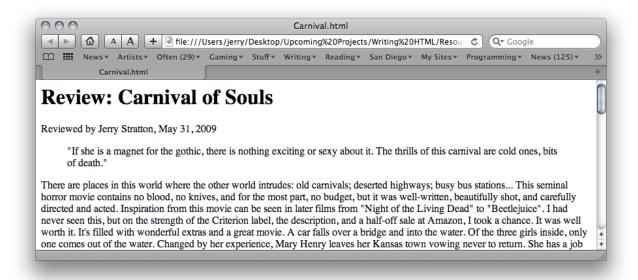
If you go and view the page in your browser, it's still just a mass of text. We've told the browser where the document is, but we still haven't given it any structure. Let's set up one headline, one paragraph, and one quote. These are the first three lines of the review.

<h1>Review: Carnival of Souls</h1>

Reviewed by Jerry Stratton, May 31, 2009

<blockquote>"If she is a magnet for the gothic, there is nothing exciting or sexy about it. The thrills of this carnival are cold ones, bits of death."</blockquote>

The tag name for the headline is "h1". The tag name for a generic paragraph is "p". And the tag name for a section of quoting is "blockquote".



It's already beginning to look quite a bit better. The HTML tags tell the browser what each bit of text is; the browser then displays that text more appropriately. Take a good look, for example, at the *blockquote* section. In the text, just as in the image above, the blockquote is on two lines. But they're not the same two lines. The web browser is ignoring the line breaks in the document; if you left in all of the extra spaces, the web browser ignores those, too. Line breaks and spaces are called "white space", and web browsers are required to collapse all white space so that it doesn't matter where the original document has line breaks, where it has indentation, or where there are multiple spaces in a row.

The only thing the web browser cares about is the location of the HTML tags.

The HEAD of the document

Human eyeballs aren't the only things reading your web page. Computers also visit your web page. In fact, computers are the only things that ever visit your web page: computers visit your web page and then store your web page in a database or display your web page to a human visitor.

Your web document has a <head> just as it has a <body>. The <head> is where you store information for computers, so that they can categorize your document and summarize it.

Title

For example, try bookmarking your web page. Most likely, the bookmark will be nothing more than the file name. That's because you haven't told computers what title they should use for your page.

At the top of your document, between the "<html>" and the "<body>", add:

```
<head>
     <title>Review: Carnival of Souls</title>
</head>
```

Save the document, and then try bookmarking it. You'll see that the title of your bookmark—most likely—is now "Review: Carnival of Souls".

You should keep your title short and descriptive. It will be used by visual browsers to title bookmarks and tabs, and by search engines to title search results. Many browsers will also show the title at the top of the browser window.

Description

You can also add a *description* to your web page. The description provides a summary of your page for search engines and other software to use. The description is contained in a *meta* tag. Meta tags come in two parts: the name and the content. They always go in the <head> of the document.

Add this meta tag to the <head> area, below the title:

<meta name="description" content="This seminal horror movie contains no blood, no knives, and for the most part, no budget, but it was well-written, beautifully shot, and carefully directed and acted." />

Notice the close of this tag. There is no "</meta>" tag, because the meta tag doesn't surround any text. In XHTML, all tags have to end, so this abbreviated form exists for tags that don't really need two parts.

In HTML, the meta tag doesn't use "/>" to end, nor does it use "</meta>". It just ends at the ">".

The meta tag also illustrates another feature of HTML tags: they can contain attributes. Attributes are in the form "name=value". Here, *description* and *name* are both attributes of this meta tag.

Keywords

It is also useful to add keywords to your web page. Keywords help software categorize your web page. Some search engines will use your keywords, although because they can be easily spammed most search engines will not provide high ranking to them. (Google, for example, claims not to use them at all.) They're still useful for internal search engines, for other software that accesses web pages, and to help you categorize your page content.

<meta name="keywords" content="Criterion, eerie, horror, influential, Herk Harvey, John Clifford" />
Keywords are listed separated by commas.

Style sheet

Often, you'll have a company-wide or site-wide style sheet that you'll want applied to all web pages. In HTML, this is "CSS" or "Cascading Style Sheets". CSS is another entire tutorial, but I do have a style sheet ready for use with this review. Add this tag to the *head* of the document:

k rel="StyleSheet" type="text/css" media="all" href="review.css" />

Like the *meta* tag, the *link* tag does not surround text, so it uses the abbreviated ending form in XHTML and has no ending in HTML. It also contains the attributes *rel*, *type*, *media*, and *href*. The first attribute, *rel*, is the *relation* between this link and the page that contains the link tag. It's a style sheet for this page. It's *type* is that it's a text file that contains CSS. It is meant for *all media* (print and screen being two common media), and it links to the *href*, or *hypertext reference*, "review.css". I'll talk more about that later, under linking.



Once you add the style sheet, reload the page. You should see the headline, first paragraph, and blockquote are now centered. The paragraph and blockquote are emphasized. And there's a horizontal line underneath the quote. These changes are all in the style sheet, and can be changed by changing the style sheet.

You can have as many <link> tags in your page's <head> as you need.

Paragraphs and Blocks

When you're marking up entire sections of your document, you basically have two kinds of tags: paragraph-level tags, and block-level tags. The main difference is that paragraph-level tags cannot contain block- or paragraph-level tags, but block-level tags can. You can't put a paragraph inside a headline, or a headline inside a paragraph. But you can put both headlines and paragraphs inside of blockquotes. Often, you have to: while paragraph-level tags can contain text, some block-level tags can only contain other tags.

The <blockquote> tag is one of these. That's why the movie quote in the <blockquote> tag is also surrounded by a paragraph tag.

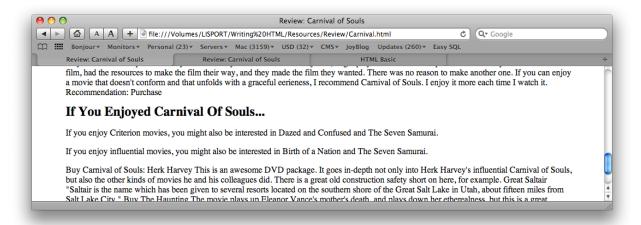
Headlines

You have six levels of headlines. Usually, you'll only use one to three of them. The others are "h2", "h3", and so on. You can think of your headline tags as outline tags. If this document were presented as an outline, what would the outline's headlines be, and what level would they be at?

We have one more headline in our document. If you look towards the bottom, you'll see "If you enjoyed Carnival of Souls...". Put an <h2> around that.

While you're at it, put paragraph tags around the two succeeding paragraphs that each begin with "If you enjoy...". I apologize for the number of them, but we'll need them to offset the images later in the tutorial.

If you reload the page, you'll see that the "if you enjoy" section is now headlined.



Remember that headline tags are for headlines, not for making large text. Headline tags mean that the marked text is the headline for the following text. If all you want to do is make some text bigger, you'll use styles to do this.

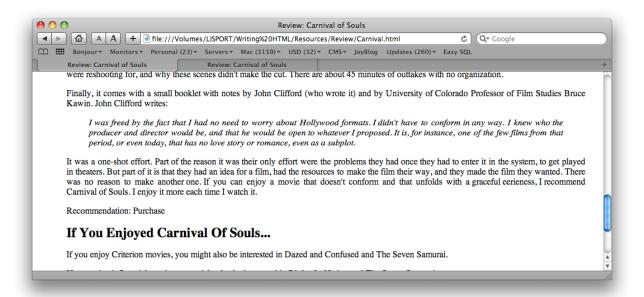
Finish the paragraphs

Just to get some practice in with paragraphs and blockquotes, let's finish up all of the rest of the paragraphs and blockquotes. Each of the blocks of text from "There are places in this world..." down to "Recommendation: Purchase" need to be placed in tags.

When you've done that, we have one blockquote in the main review. The third-to-last paragraph, beginning with "I was freed by the fact that I had no need..." should have a blockquote tag around its paragraph tag.

```
The DVD set also contains a thirty-minute documentary, "The Movie That Wouldn't Die" covering, a revival showing of the film in Lawrence Kansas where it was mostly filmed. The showing brought together directory Harvey, writer
66
67
        John Clifford, and stars Candace Hilligoss and Sidney Berger. It's a fun watch.
        There are written interviews with Herk Harvey, John Clifford, and Candace Hilligoss. The writing is interspersed with photos and artwork. Candace Hilligoss's interview ends with closeups of the movie poster showing the incredible artwork. It's the same poster used for the cover of the DVD box, but the cover is too small to do it
69
70
71
72
73
74
75
76
77
78
80
81
82
        justice. it seems a bit weird reading text on the television, but they're good interviews.
         There are a lot of outtakes, different takes on scenes. These would be a lot more interesting if there were a
         commentary talking about what they were reshooting for, and why these scenes didn't make the cut. There are about
        45 minutes of outtakes with no organization. 
        Finally, it comes with a small booklet with notes by John Clifford (who wrote it) and by University of Colorado
Professor of Film Studies Bruce Kawin. John Clifford writes:
           83
84
           instance, one of the few films from that period, or even today, that has no love story or romance, even as a
           subplot.</blockauote>
85
86
87
         It was a one-shot effort. Part of the reason it was their only effort were the problems they had once they had to
        enter it in the system, to get played in theaters. But part of it is that they had an idea for a film, had the resources to make the film their way, and they made the film they wanted. There was no reason to make another one. If you can enjoy a movie that doesn't conform and that unfolds with a graceful eerieness, I recommend Carnival of
89
90
91
        Souls. I enjoy it more each time I watch it.
92
93
                                                                                                                               Recommendation: Purchase
    <h2>If you enjoyed Carnival of Souls...</h2>
           antf you aming Chitanian magice, you might also be interpreted in Based and Canfiged and The Course Company /so
```

Once you've done that, the main part of the review should be readable, if bland.



Lists

Some collections of text are lists. There are several kinds of lists within HTML. In this review, there's a "definition list" at the bottom of the page. A definition list is a list of items where each item consists of a *title* and a *description*. Look at the remaining unmarked text at the bottom of the review. It looks very much like a dictionary entry; there are four items in the list and each item has a short title and a longer description.

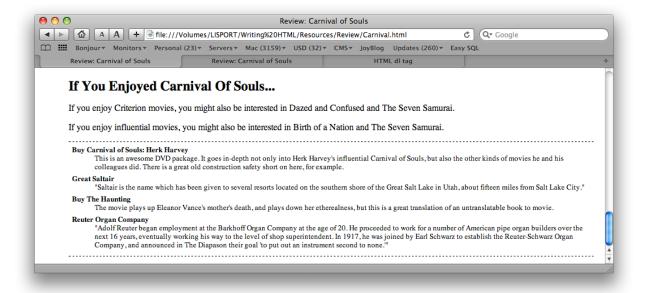
1. Surround that entire section with <dl> and </dl>.

- 2. Surround each title with <dt> and </dt>.
- 3. Surround each description with <dd> and </dd>.

It should look like this:

```
100 <dl>
101
        <dt>Buy Carnival of Souls: Herk Harvey</dt>
102
                 <dd>This is an awesome DVD package. It goes in-depth not only into Herk Harvey's influential Carnival of
103
                 Souls, but also the other kinds of movies he and his colleagues did. There is a great old construction
104
                 safety short on here, for example.</dd>
105
106
        <dt>Great Saltair</dt>
                 <dd>"Saltair is the name which has been given to several resorts located on the southern shore of the Great
107
108
                 Salt Lake in Utah, about fifteen miles from Salt Lake City."</dd>
109
110
        <dt>Buy The Haunting</dt>
                 <dd>The movie plays up Eleanor Vance's mother's death, and plays down her etherealness, but this is a great
111
112
                 translation of an untranslatable book to movie.</dd>
113
114
        <dt>Reuter Organ Company</dt>
115
                <dd>"Adolf Reuter began employment at the Barkhoff Organ Company at the age of 20. He proceeded to work for a
                 number of American pipe organ builders over the next 16 years, eventually working his way to the level of shop superintendent. In 1917, he was joined by Earl Schwarz to establish the Reuter-Schwarz Organ Company,
116
117
118
                 and announced in The Diapason their goal 'to put out an instrument second to none.'"</dd>
```

Reload the page and you'll see that the jumbled text at the bottom is now structured as a list of titles and descriptions.



The style sheet has also added dashed lines around the list of other sources.

Character styles

So far, every tag we've looked at has contained no text, or it has contained entire paragraphs (or, in the case of *blockquote*, potentially has contained multiple paragraphs).

Emphasis

You'll often want to emphasize some text within the web page. In visual browsers, emphasis is usually displayed using *italics* or **bold**. In HTML, we'll use (for normal emphasis) and (for stronger emphasis).

In the first paragraph of the review, put tags around blood, knives, and budget:

This seminal horror movie contains no blood, no knives, and for the most part, no budget, but it was

Reload the web page, and you'll see that those words are now emphasized using italics.

Now, what's the difference between this and the italics for, say, the blockquotes? The difference is in meaning. The blockquotes are italicized because we want to set the blockquotes off from the rest of the text. The emphasized text is italicized because we want that text to be emphasized to the reader. If the web page were being displayed in a non-visual way, we wouldn't want the blockquotes to be italicized; it wouldn't make sense. But we would still want the emphasized text to be emphasized, whatever that means in the non-visual browser displaying it.

For stronger emphasis, go to the sixth paragraph of the review, and put tags around *original* theatrical version and 1989 restoration.

The DVD set contains both the original theatrical version, which had been cut without the director's input, and the director's 1989 restoration; the latter brings the original 78 minutes to 83 minutes. Each version is on a

Reload the web page, and you'll see that those phrases are emphasized using bolded text.

Links

We've already done a little bit of linking. The kas linked to the style sheet for the web page. That link is invisible (except for its effects). We can also create links in the main text of the web page that the visitor can use to visit other web pages. For these kinds of links, we use the <a>, or anchor, tag. This is one of the most important features of the web: the ability to immediately cross-reference between different pages at different sites.

There are two phrases ripe for linking. "It keeps coming back, though" in the eighth paragraph can be linked to the Saltair web site, and "Reuter Organ Company" in the next paragraph can be linked to that company's web site.

```
<a href="http://www.thesaltair.com/">It keeps coming back, though.</a><a href="http://www.reuterorgan.com/">Reuter Organ Company</a>
```

Save and reload the page, and you can now click on those phrases to follow through to those web sites.

The href attribute of an a tag (or a link tag) is a URL. For offsite links, it always begins with "http://"; it is the link that you see in the URL bar of your browser after you visit the page. Sometimes you can use abbreviated URLs to get to a page, such as leaving out the .com or the http://, but in an href attribute, you have to use the full URL.

Go ahead and link the titles in the list at the bottom of the page:

Carnival of Souls	http://en.wikipedia.org/wiki/Carnival_of_Souls
	http://www.thesaltair.com/
The Haunting	http://en.wikipedia.org/wiki/The_Haunting_(1963_film)
Reuter Organ Company	http://www.reuterorgan.com/

If you don't want to type the URLs, do a web search for them. This is always a good idea when adding links to your page: make sure the links are current, and use copy-and-paste to reduce the possibility of typos.

There are two other ways of linking that can be used for local files: files that exist on the same server as the web page. We've already used one of them. If the file you're linking to is in the same folder as the web page, you can just use the file name. For example, when linking to the style sheet we just used the filename, "review.css". You might also store some types of files in a folder that's in the same folder as the web page, and then you would put the folder name in front of the file name. For example, "images/saltair.jpg".

If the file you're linking to is on the same server but in a different part of the site, you can use the "full path". For example, it isn't uncommon to have a library area on your site that contains scripts, style sheets, and images. If our style sheet were in that library, the href might be "/library/css/review.css". By beginning the href with a forward slash, the browser knows to use this server, but start at the beginning.

Always use the forward slash to separate parts of URLs, as in the above examples.

Validation

I've talked a little about the difference between XHTML and HTML. But how does the browser know the difference? You need to tell it.

Document Type

Unless you've got a good reason, I recommend using either XHTML 1.0 Strict or HTML 4.0.1 Strict. In this tutorial I'm using XHTML. You can read more about them on Wikipedia and at W3Schools.com.

Replace the opening httml tag with this:

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">

You may find it useful to copy and paste, and I've also provided this snippet in the snippets folder of the resources folder.

You must specify a DOCTYPE. If you don't, browsers are likely to assume you don't know what you're doing, and not let your page use some features of modern HTML and XHTML.

Character set

The computer world is filled with different character sets, and in a world where lots of different software on lots of different computers all read the same files, it's important to know which character set was used. Computers work with numbers; every letter and every digit and everything that comes off of your typewriter is represented as a number when it gets saved to a file. The problem is that different software often use different numbers to represent these characters. For example, character 211 in one character might be the closing double quotes. In another, 211 could be an accented capital O.

It's a good idea to specify which character set you're using. I recommend utf-8 unless you've got a good reason to do otherwise.

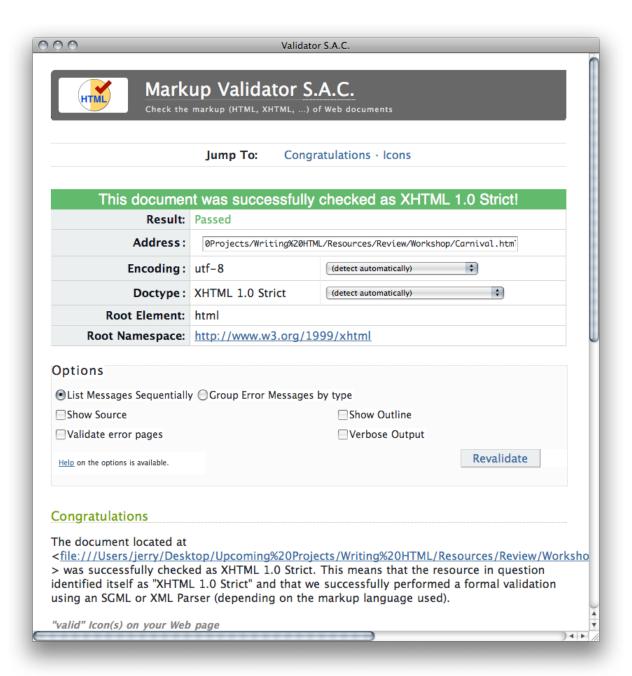
<meta http-equiv="Content-Type" content="text/html;charset=utf-8" />

This is a standard <meta> tag and it goes in the <head> of the document. You'll also need to tell your text browser that you're using UTF-8.

Validation

You can (and should, often) go to http://validator.w3.org/ to validate your web pages. You can also download a stand-alone application for Mac OS X at http://habilis.net/validator-sac/.

When you validate your web page, the validator will tell you if there are any mistakes in the actual HTML or in the character set. Fixing these errors will make your web page display more reliably on more web browsers.



Flair

Style Sheets

The purpose of HTML is to structure your document. Resist the urge to use it for layout. Layout is device-specific, but structure is universal. A well-structured document is easier for new, unanticipated technologies to read, easier for you to maintain, and easier for you to vary the layout on.

When it comes time to vary the visual layout, you have style sheets (CSS). We're not going to cover style sheets in this tutorial, but we will cover how to work with them.

Classes and IDs

Each tag can have a class and it can have an ID. An ID uniquely identifies that tag among all other tags on the page. An ID can only appear once throughout the web page.

```
<dl id="ratings">
```

IDs can be used in CSS and in JavaScript to reference that specific tag.

A class can appear as many times as it is needed. It is usually used only in CSS.

```
<a class="hot">
```

<u>Styles</u>

You can also add the *style* attribute to any display tag. This allows you to put CSS directly into the HTML.

```
<hl style="background-color: green;">
```

Normally, you'll want to avoid this: it's easier to change layouts later if the style information is in a separate file.

DIVs and SPANs

Sometimes you need to section off a portion of a web page or paragraph for special styling, but there is no appropriate tag to use. When this happens, you can use a <div> tag or a tag. The <div> tag is a block-level tag that can contain any other block-level or paragraph level tag (including other div tags). The tag is a character-level tag that can contain text and other character-level tags, but not paragraph-level or block tags.

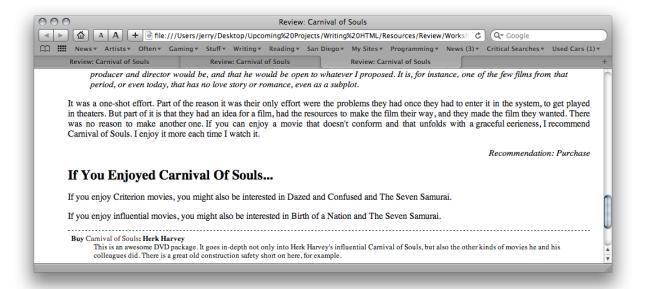
You'll almost always give div and span tags either a class or an ID (or both), because there isn't any other reason to have them in a document.

Put a div around the main part of the review. Give it the ID of "review". It should begin after the "There are places" paragraph but before the "I had never seen" paragraph. It should end between the "It was a one-shot" and "Recommendation: Purchase".

<div id="review">

I had never seen this, but on the strength of the Criterion label, the description, and a half-off sale at Amazon, I took a chance. It was well worth it. It's filled with wonderful extras and a great movie.

Reload the web page, and the recommendation to purchase that appears *after* this div tag closes will now become italicized and align itself to the right. The style sheet specifies that any paragraph that comes immediately after the review must be styled in that manner.



Images

Our web page is looking a lot more like a web page now than it did when we started, but no movie review would be complete without stills from the movie. There are four images in the workshop folder, and we'll add each of them to the review. The first image will be a screen capture of the main character looking at her reflection. Directly underneath the "<div id="review">", add the image and a paragraph to caption it:

Does Mary have no soul, or one too many?

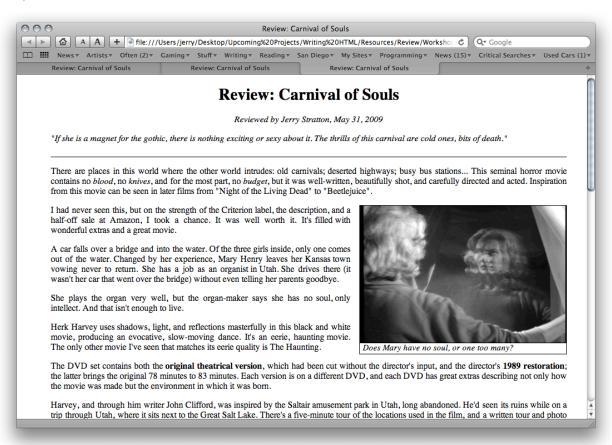
For our purposes, the "src" attribute is exactly like the "href" attribute we've already seen. It can contain URLs and local references in exactly the same way. Here, it refers to the file called "Reflections.jpg" in the same folder as the web page.

Alternate text

A picture may be worth a thousand words, but words are much more versatile, especially on the Internet. Computer software—such as search engines and alternative browsers—find it much easier to reposition text to different viewing mechanisms than they do images. All images should contain an alt attribute, even if it's empty. An empty alt attribute tells alternative browsers that if they can't display this image, they shouldn't display anything.

Pullout

We'd like to combine these images with these captions. Sectioning off an img tag and a paragraph tag? That's a job for a div. The style sheet already has a class ready for image pull-outs, called "pull". Surround the image and the paragraph with a div that has a class of "pull".



The "pull" class is defined in the style sheet; take a look at it if you're interested.

Now let's do the next three. These images, their alt text, and their captions, are in the "images" folder of the resources folder, if you find that easier to copy from.

The Saltair image goes in front of the paragraph that begins "One of the problems with the site is that the Great Salt Lake has a very dynamic water level".

```
<div class="pull">
        <img src="Pavilion.jpg" alt="Saltair Pavilion" />
        The Saltair ruins were the inspiration for the film.
</div>
```

The Star34 image goes in front of the paragraph that begins "There's a commentary on the restored version".

And the diving Mule image goes in front of the paragraph that begins "There are a lot of outtakes".



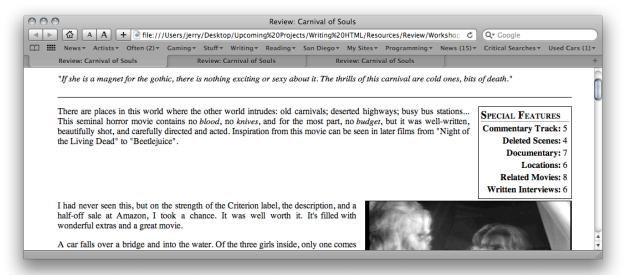
The images alternate between the left and right because the style sheet tells them to.

Tables

Another thing that might be useful for readers is a couple of tables of information about the DVD and the movie. Let's add a list of DVD features. (This table, and the other tables we'll create, are in the "tables" folder of the resources folder.) Add this below the movie quote at the top of the page, and above the "There are places in this world" paragraph.

```
<div id="features">
    <h3>Special features</h3>

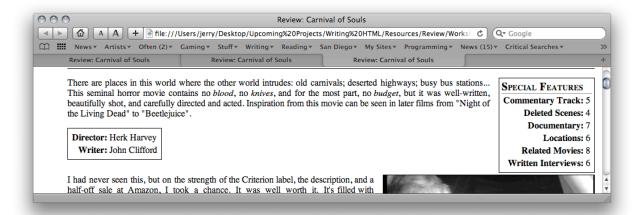
    Commentary Track5
```



This is the most complex tag we've seen (and are likely to see). Like the definition list tags, the table tag contains a series of other tags. Tables contain rows, and rows contains cells. The
 try tag marks table rows. The
 try tag marks table headers, and the tag marks table data. You can have as many data cells in a row as you need. You'll probably want only one header cell, however. Here, we have the header cells all in the left column; another common format is for all of the header cells to be in the top row.

Let's try another, simpler table. After the "There are places" paragraph but before the div with id "review", add:

```
    DirectorHerk Harvey
    WriterJohn Clifford
```



You should be careful with tables. They're so simple to create, you can easily make a web page that can't be read by anyone but you. Remember that there are lots of different web browsers out there, some of them that don't even use computer screens. Simpler is almost always better!

Lists

If you pulled the director and writer out of review.txt, you'll notice that there was a third row I didn't use. That row contained a bulleted list. Go ahead and add it to the movie table as a third row:

```
Formats

* Academy Ratio

* 1.92 Widescreen
```

Save this and reload it in your web browser, and it all runs together. Browsers ignore white space. If we want the browser to treat those two lines as items in a list, we need to tell it that they are two items in a list.

Things that would normally be displayed as unnumbered lists are *unordered* lists. The tag for unordered lists is The tag for each item in the list is .

The defining characteristic of unordered lists are that the order doesn't matter; you would never use numbers to count up an unordered list; if you needed a marker, you would use a bullet for each item. We don't have any in this document, but if the list is *ordered*, the tag is
 List items remain marked by Ordered lists are usually displayed with numbers, and sometimes with letters, such as 'a', 'b', 'c', and so on.

List items are both paragraph-level and block-level tags. The tag can contain most paragraph-level and block-level tags. The
 tags can only contain tags.

Let's take a look at one more list. This one is also in a table, so add the table first. After the recommendation paragraph and before the "If you enjoyed" headline, create a table with these three rows:

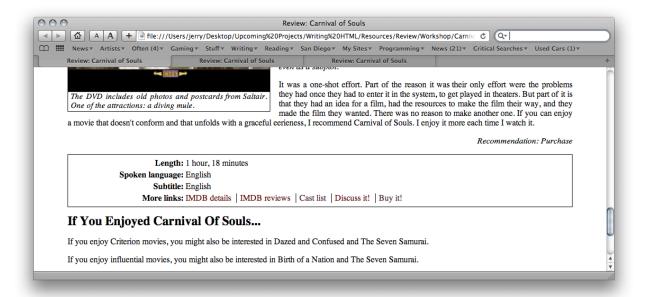
This list has a class applied to it, and the class corresponds to a set of styles in the style sheet that drastically alter how the list appears in visual browsers. Underneath, however, it remains a list and is treated that way by software.

More about URLs

That list of links we added in the previous section aren't actually linked. Here are the links for each item. Add them with an "a" tag.

	http://www.imdb.com/title/tt0055830/
IMDB reviews	http://www.imdb.com/title/tt0055830/externalreviews
	http://www.imdb.com/title/tt0055830/fullcredits
	http://groups.google.com/groups/search?as_q=Carnival%20of%20Souls&as_ugroup=*movies*
Buy it!	http://www.amazon.com/exec/obidos/ISBN=1559409002/

The first three URLs look normal; they're off-site links for more information about the movie. The fourth URL, for "Discuss it!" is a form submission. Everything after the question mark is sent to a computer program on that server; the computer program decides what to send back based on that query. Each part of the query is separated by an ampersand (& amp;). Special characters, such as spaces, need to be specially encoded, because special characters aren't allowed in URLs. Often, these queries result in the server "querying" a database and returning the results of that query. (Here, it queries their database of discussion group postings.)



Deleted and inserted text

If you're writing news stories or blog postings, you'll often amend your text later: you'll discover that something you wrote was incorrect or misleading, and you'll want to correct your text. But it's bad form to change a news posting or a blog posting without warning; these types of pages are assumed to exist at a specific moment in time. HTML has a special tag for marking text that's been "removed" as well as for marking the new text that replaces it. These are the and <ins> tags.

Surround text that needs to be "deleted" with the tag. Surround new text with the <ins> tag. Visual browses will usually mark deleted text by striking through it, and new text by underlining it.

Both of these tags allow two special attributes: the date and time of the change (datetime) and a URL to a web page explaining the change (cite). Neither of these attributes are required, although they obviously add greater precision. The "cite" attribute takes a full or partial URL just like the "href" and "src" attributes do.

The "datetime" attribute is a bit hard to read. The format is "YYYY-MM-DD" for the date portion, "hh:mm:ss" for the time portion, and either "Z" (universal time) or "+/-hh:mm" for the hours and minutes ahead of/behind universal time. All three sections are required; the date and the time are separated by a "T". The datetime attribute as I write this is "2009-09-08T17:22:51-07:00" or "2009-09-08T00:22:51Z". The hours, minutes and seconds are all required; however, if they are unknown each may be left at "00". For example, "2009-09-08T00:00:00Z" if you only know the date, and "2009-09-08T00:22:00Z" if you don't know the seconds.

You can use these tags to as if they were block-level tags, or you can use them as if they were character-level tags. Thus, you can mark text within a paragraph as deleted (or inserted), or you can mark a series of blocks or paragraphs as deleted or inserted.

For example:

```
>
       <del cite="http://www.hypocritae.com/?ART=6" datetime="1994-04-20T19:08:44-0400">
           Whatever goes upon two legs is an enemy.
   <1i>
       <del cite="http://www.hypocritae.com/?ART=6" datetime="1994-05-20T19:08:44-0400">
           Whatever goes upon four legs, or has wings, is a friend.
       </del>
   i>
       <del cite="http://www.hypocritae.com/?ART=6" datetime="1994-06-20T19:08:44-0400">
           No animal shall wear clothes.
       </del>
   li>
       <del cite="http://www.hypocritae.com/?ART=6" datetime="1994-07-20T19:08:44-0400">
           No animal shall sleep in a bed.
       </del>
   <del cite="http://www.hypocritae.com/?ART=6" datetime="1994-08-20T19:08:44-0400">
           No animal shall drink alcohol.
       </del>
   <del cite="http://www.hypocritae.com/?ART=6" datetime="1994-09-20T19:08:44-0400">
           No animal shall kill any other animal.
       </del>
   li>
       All animals are equal.
       <ins cite="http://www.hypocritae.com/?ART=6" datetime="1994-09-20T19:08:44-0400">
           But some animals are more equal than others.
       </ins>
```

Entities (special characters)

In the past, any special characters needed to be specially encoded. It isn't as necessary today, as long as you specify a character set and make sure that you always stick with that character set. But entity codes can still be useful if you're working with multiple data sources of unknown character set, or you have multiple people working on the same document and you're worried that they won't all use the same character set.

Ellipses and m-dashes

Two common special characters are m-dashes and ellipses. There are two ellipses in this document and one m-dash. The ellipses are currently three periods, and the m-dash is two smaller dashes. Change them to … and —, respectively.

Quotes

The most common special characters you'll use are typographer's quotes, since they improve the readability of your text.

Left double quote		"
Right double quote	"	"
Left single quote		'
Right single quote	,	'

You may find the search and replace feature of your text editor useful for this, but go ahead and change all of the straight quotes to the appropriate typographer's quote. Hint: most single quotes are right single quotes, since they're used for contractions. Warning: make sure you *don't* change the straight quotes used inside HTML tags to mark attribute values!

In grade school, I used to look forward to those " educational" films about faraway places or road safety, good or bad. Some of the good ones might have been made by Herk Harvey. Criterion includes several " educational" films directed by Herk for Centron, and about four of his commercial films to give you an idea of what he was doing before and after " Carnival of Souls". They range from " Signals: Read' em or Weep" (my favorite) to promos for Korea, Jamaica, and Kansas itself where Centron was located. The Kansas promo (Star 34, after Kansas's star on the U.S. flag) is especially interesting because it shows a very young Herk Harvey, when he had just started working for the Centron Corporation.

Other characters

Other commonly encoded characters are accented letters and other diacritics.

Letter	′ accent	` accent	" umlaut	^ circumflex
a	á	à	ä	â
e	é	è	ë	ê
i	í	ì	ï	î
0	ó	ò	ö	ô
u	ú	ù	ü	û

You can capitalize the first letter to get the capitalized version of that letter.

Others include "ç" for ç and "ñ" for ñ.

Ampersands

Because special characters are encoded using the ampersand, the one special character you do always have to worry about is the ampersand. You should never have a "bare" ampersand in your web pages. All ampersands—even the ones in URLs—must be encoded using "& amp;".

Extra Credit

Special tags

There are other tags than these; any good XHTML/HTML book or web site will describe them. Commonly-used ones are <sup> and <sub> for superscript and subscript, <cite> for citations, and
 for line breaks. The line break tag doesn't surround any text, so it appears as
 in XHTML and
 in HTML.

Meaningless markup

Sometimes you're marking up printed documents and you need to use print formatting, solely for replicating print formatting and not to impart any meaning. The underline tag is <u>, the italics tag is <i>, and the bold tag is . There's also a <strike> tag for striking a line through text. Normally, though, you'll use the meaningful tags: if you want to emphasize text, use or . If you want to mark text as having been rescinded or deleted, use .

Note that the <strike> tag and the <u> tag are only available in HTML. In XHTML, you'll underline and strike out text using style sheets if you need to avoid the meaningful tags.

Why XHTML 1.0?

The two real choices for HTML as I write this are HTML 4.0.1 and XHTML 1.0. Why do I use XHTML? Because XHTML is XML, and the tools for working with XML are more robust and reliable, in my experience, than the tools for working with HTML. When I write a program to take a full web page and return a snippet, I can safely break up the web page based on the XHTML tags starting and ending reliably. This isn't something I can do with HTML.

When I validate an XHTML web page, I know that the tags begin and end. When I validate an HTML web page, I don't: HTML validates fine without closing tags. The closing tags for paragraphs and the various list items can be there, but they don't have to be there. This means that validation won't tell me that they're missing. That can make it easier to create web pages using HTML, but it makes it a lot harder to get information out of those pages. It can also cause problems with white space.

I also use XML a lot for passing data back and forth between web applications. So it's easier for me to think in XHTML.

XHTML 1.1 and XHTML 1.2 appear to have been designed for special purposes and not for general use on the web.

HTML 4.0.1

You can see an example of HTML 4.0.1 in the Samples folder, as "Carnival (HTML).html". This is the same as the final "Carnival (Completed).html" but modified to conform to HTML 4.0.1. If you view both of those files side-by-side in your browser, they should display the same (I tested it in Safari 4 and Firefox 3.5).

HTML is much more free-form than XHTML. Many tags only need to be marked at their opening and don't need an ending. You can also mix capitalization to make some tags stand out more.

In this page, I've removed the abbreviated closing slash from the meta tags and the link tag, because they're invalid in HTML 4; I also removed the abbreviated closing slash from the img tags because they generate a warning.

I've removed the ending tags for all of the paragraphs and list items, and all of the table rows and most of the table cells. Two table cells still needed them: the header cell for "More links" and the header cell for "Formats". That's because without the closing tag, browsers will include the line breaks and other white space as part of the cell data, resulting in an extra space on those right-aligned cells.

If you find that HTML 4.0.1 works better for you, there's nothing wrong with using it. This is the doctype that goes at the very top of the file:

<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN" "http://www.w3.org/TR/html4/strict.dtd">
It uses a simple <html> tag to open the web page.

Blog comments

One common use for HTML even for people who don't write web pages is for marking up blog comments. If you're commenting in a blog, and the blog says that it supports HTML, it will let you use HTML to emphasis text, add links, and sometimes even add images. Usually, this will mean the emphasis tags, the meaningless markup tags, and links.

If you're writing a blog comment, and it lets you use link tags, it's a lot nicer to make a link with the real URL, rather than making a compressed URL as is often done to save space. It helps people who are careful where they click see where the URL is really going.

More Information

Links

The most useful book I've seen for HTML is HTML & XHTML: The Definitive Guide, from O'Reilly. Nowadays I tend to use the web more often, however. Some of the web pages I've used while writing this tutorial are the w3schools HTML and XHTML pages.

HTML: The Definitive Guide	http://www.hoboes.com/Mimsy/hacks/html-xhtml-definitive/
	http://tuppis.com/smultron/
	http://www.hoboes.com/Mimsy/hacks/cascading-style-sheets/
	http://www.w3schools.com/html/
	http://www.w3schools.com/xhtml/
	http://validator.w3.org/
Standalone Markup Validator for OS X	http://habilis.net/validator-sac/

"The best book on programming for the layman is Alice in Wonderland; but that's because it's the best book on anything for the layman."

GNU Free Documentation License

Version 1.1, March 2000

Copyright (C) 2000 Free Software Foundation, Inc. 59 Temple Place, Suite 330, Boston, MA 02111-1307 USA Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

0. Preamble

The purpose of this License is to make a manual, textbook, or other written document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. Applicability and Definitions

This License applies to any manual or other work that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you".

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (For example, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them. The "Invariant Sections" are certain Secondary Sections whose titles are

The "Invariant Sections" are certain Secondary Sections whose titles ar designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, whose contents can be viewed and edited directly and

straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup has been designed to thwart or discourage subsequent modification by readers is not Transparent. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML designed for human modification. Opaque formats include PostScript, PDF, proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

2. Verbatim Copying

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3. You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. Copying in Quantity

If you publish printed copies of the Document numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a publicly-accessible computer-network location containing a complete Transparent copy of the Document, free of added material, which the general network-using public has access to download anonymously at no charge using public-standard network protocols. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the

4. Modifications

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
- List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has less than five).
- 3. State on the Title page the name of the publisher of the Modified Version, as the publisher.
- 4. Preserve all the copyright notices of the Document.
- Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
- Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- 8. Include an unaltered copy of this License.
- 9. Preserve the section entitled "History", and its title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- 10. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
- 11. In any section entitled "Acknowledgements" or "Dedications", preserve the section's title, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
- 12. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
- Delete any section entitled "Endorsements". Such a section may not be included in the Modified Version.
- 14. Do not retitle any existing section as "Endorsements" or to conflict in title with any Invariant Section.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties —for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. Combining Documents

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections entitled "History" in the various original documents, forming one section entitled "History"; likewise combine any sections entitled "Acknowledgements", and any sections entitled "Dedications". You must delete all sections entitled "Endorsements."

6. Collections of Documents

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. Aggregation with Independent Works

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, does not as a whole count as a Modified Version of the Document, provided no compilation copyright is claimed for the compilation. Such a compilation is called an "aggregate", and this License does not apply to the other self-contained works thus compiled with the Document, on account of their being thus compiled, if they are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one quarter of the entire aggregate, the Document's Cover Texts may be placed on covers that surround only the Document within the aggregate. Otherwise they must appear on covers around the whole aggregate.

8. Translation

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License provided that you also include the original English version of this License. In case of a disagreement between the translation and the original English version of this License, the original English version will prevail.

9. Termination

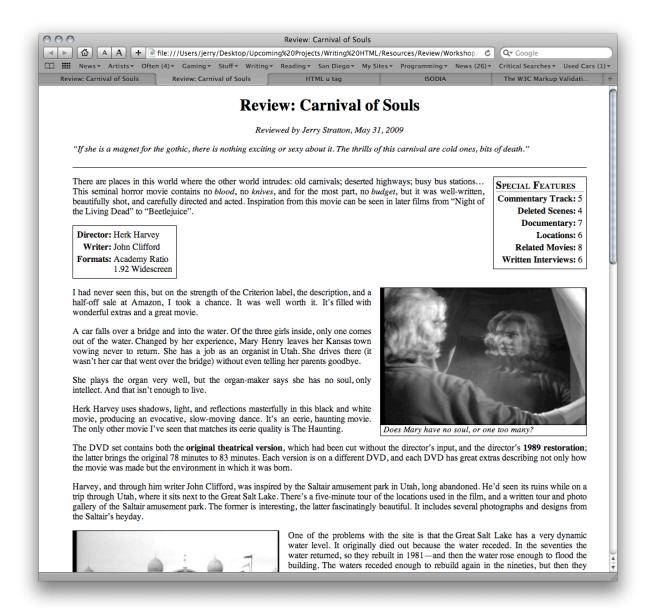
You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

10. Future Revisions of this License

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

HTML for the Web



This simple tutorial shows you how to "mark up" documents to make them readable on the web in the context of today's style sheet-oriented web pages.