LMSGI

6. HTML5

Alejandro Amat Reina

Un poco de historia

HTML 1.0–2.0

These were the early days; you could fit everything there was to know about HTML into the back of your car. Pages weren't pretty, but at least they were hypertext enabled. No one cared much about presentation, and just about everyone on the Web had their very own "home page." Even a count of the number of pencils, paperclips, and Post-it notes on your desk was considered "web content" back then (you think we're kidding).

1989 **1991**

HTML 3

The long, cold days of the "Browser Wars." Netscape and Microsoft were duking it out for control of the world. After all, he who controls the browser controls the universe, right?

At the center of the fallout was the web developer. During the wars, an arms race emerged as each browser company kept adding their own proprietary extensions in order to stay ahead. Who could keep up? And not only that, back in those days, you had to often write two separate web pages: one for the Netscape browser and one for Internet Explorer. Not good.

1995

HTML 4

Ahhh…the end of the Browser Wars and, to our rescue, the World Wide Web Consortium (nickname: W3C). Their plan: to bring order to the universe by creating the ONE HTML "standard" to rule them all.

The key to their plan? Separate HTML's structure and presentation into two languages—a language for structure (HTML) and a language for presentation (CSS)—and convince the browser makers it was in their best interest to adopt these standards.

But did their plan work?

Uh, almost…with a few changes (see HTML 4.01).

1998

Un poco de historia

HTML 4.01

The good life: HTML 4.01 entered the scene in 1999, and was the "must have" version of HTML for the next decade.

4.01 wasn't really a big change from 4.0; just a few fixes were needed here and there. But compared to the early days of HTML (when we all had to walk barefoot in six feet of snow, uphill both ways), HTML 4.01 allowed us all to sleep well at night knowing that almost all browsers (at least the ones anyone would care about) were going to display our content just fine.

XHTML 1.0

Just as we were all getting comfortable, a shiny object distracted everyone. That shiny object was XML. In fact, it really distracted HTML, and the two got hitched in a shotgun marriage that resulted in XHTML 1.0.

XHTML promised to end all the woes of the Web with its adherence to strictness and new way of doing things.

The only problem was, most people ended up hating XHTML. They didn't want a new way to write web pages, they just wanted to improve what they already had with HTML 4.01. Web developers were far more interested in HTML's flexibility than XHTML's strictness. And, more and more, these developers wanted to spend their time creating web pages that felt more like

HTML5

Of course, with no support from the community, the marriage didn't end well and was replaced by new version of HTML named HTML5. With its support for most of the HTML 4.01 standard, and new features that reflect the way the Web has grown, HTML5 is what developers were looking for. And, with features like support for blog-like elements, new video and graphic capabilities, and a whole new set of capabilties aimed at building web applications, HTML5 was destined to become the standard.

To be honest, the divorce of HTML and XML took a lot of people by surprise, leading to confusion about what HTML5 actually *is* for a while. But that's all been sorted out, so read on to find out what HTML5 means to

1999 **2001** **2009** **2012** **????**

Alejandro Amat Reina

HTML Archaeology

We did some digging and found some old HTML 4.01 and XHTML 1.1 pages. These pages use a doctype, at the very top of the HTML file, to tell the browser which version of HTML they're using. We've snipped out a couple of doctypes for you to look at. Check them out below...

This is specifying a document type for this page to the browser.

This means that <html> is the root (first) element in your page.

This just means the HTML 4.01 standard is publicly available.

This part says we're using HTML version 4.01 and that HTML markup is written in English.

```
<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01//EN"
    "http://www.w3.org/TR/html4/strict.dtd">
```

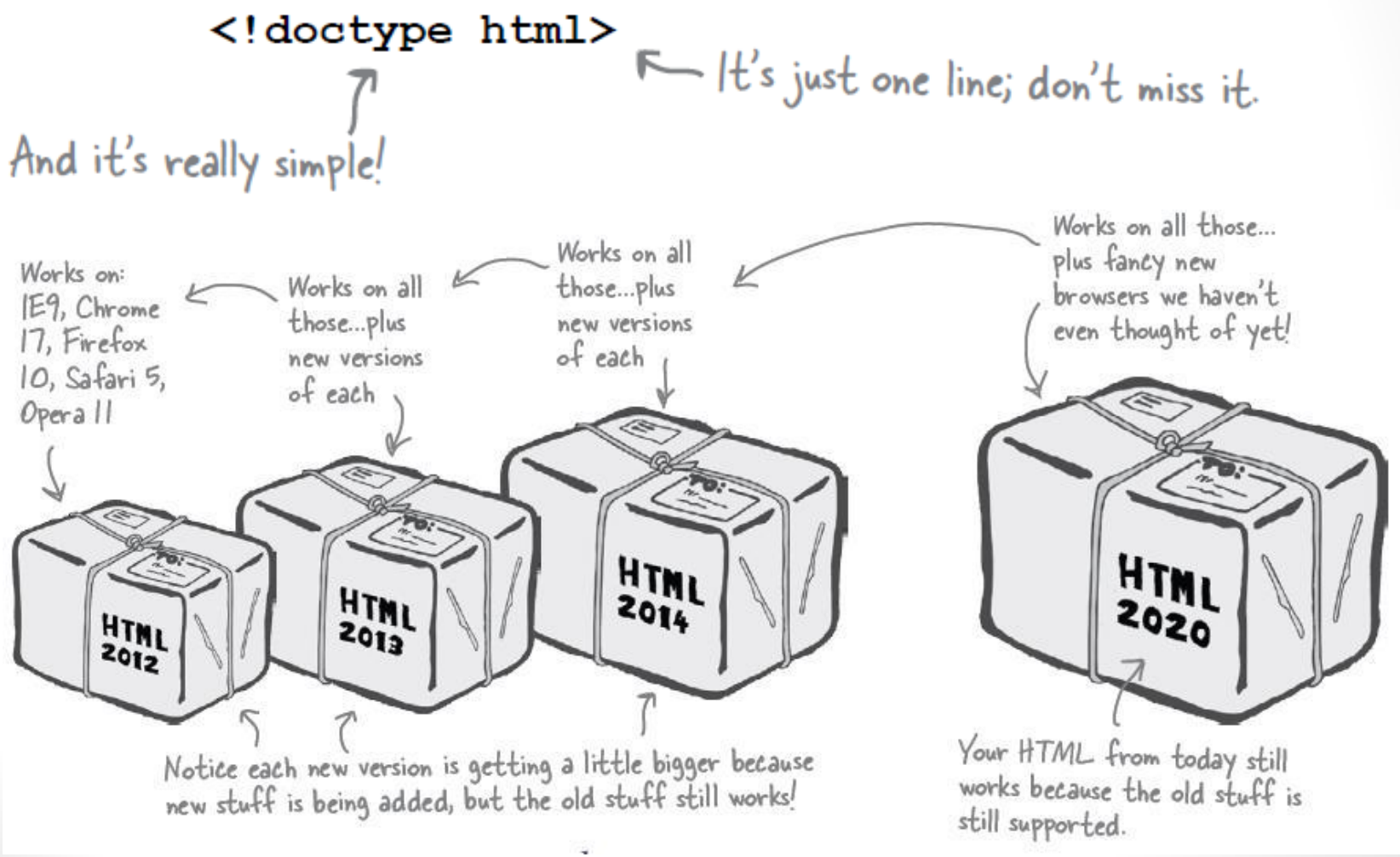
Just like the HTML DOCTYPE, this is a public document type.

It's still a version of HTML—an XML version.

It's for the XHTML 1.1 version of XHTML.

```
<!DOCTYPE html
    PUBLIC "-//W3C//DTD XHTML 1.1//EN"
    "http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">
```

HTML5 doctype



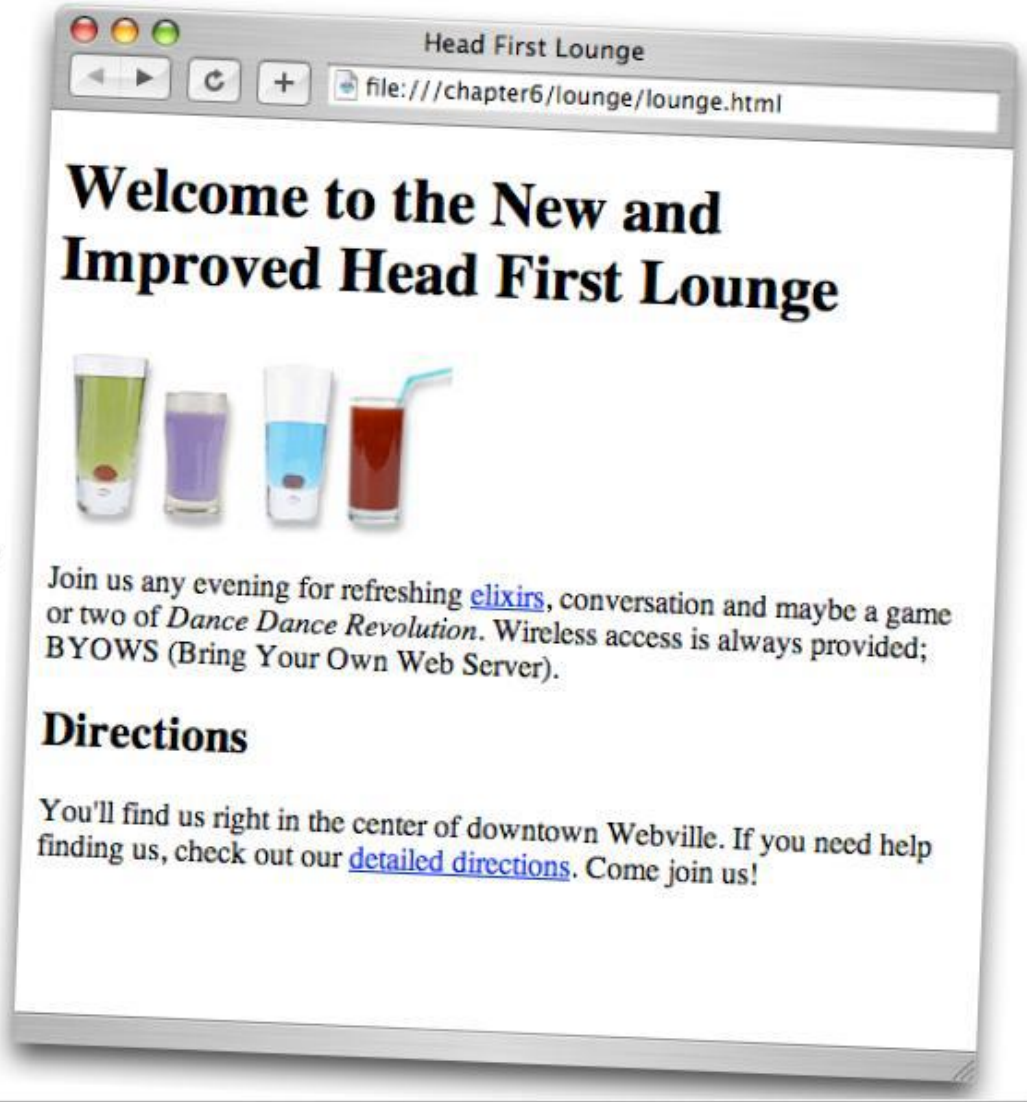
Añadiendo la definición del tipo de documento

Here's the doctype line. Just add it as the very first thing in the "lounge.html" file.

You can write DOCTYPE or doctype. Both work.

```
<!doctype html>
<html>
  <head>
    <title>Head First Lounge</title>
  </head>
  <body>
    <h1>Welcome to the New and Improved Head First Lounge</h1>
       <img src="drinks.gif">
```

Change and test drive "lounge.html"

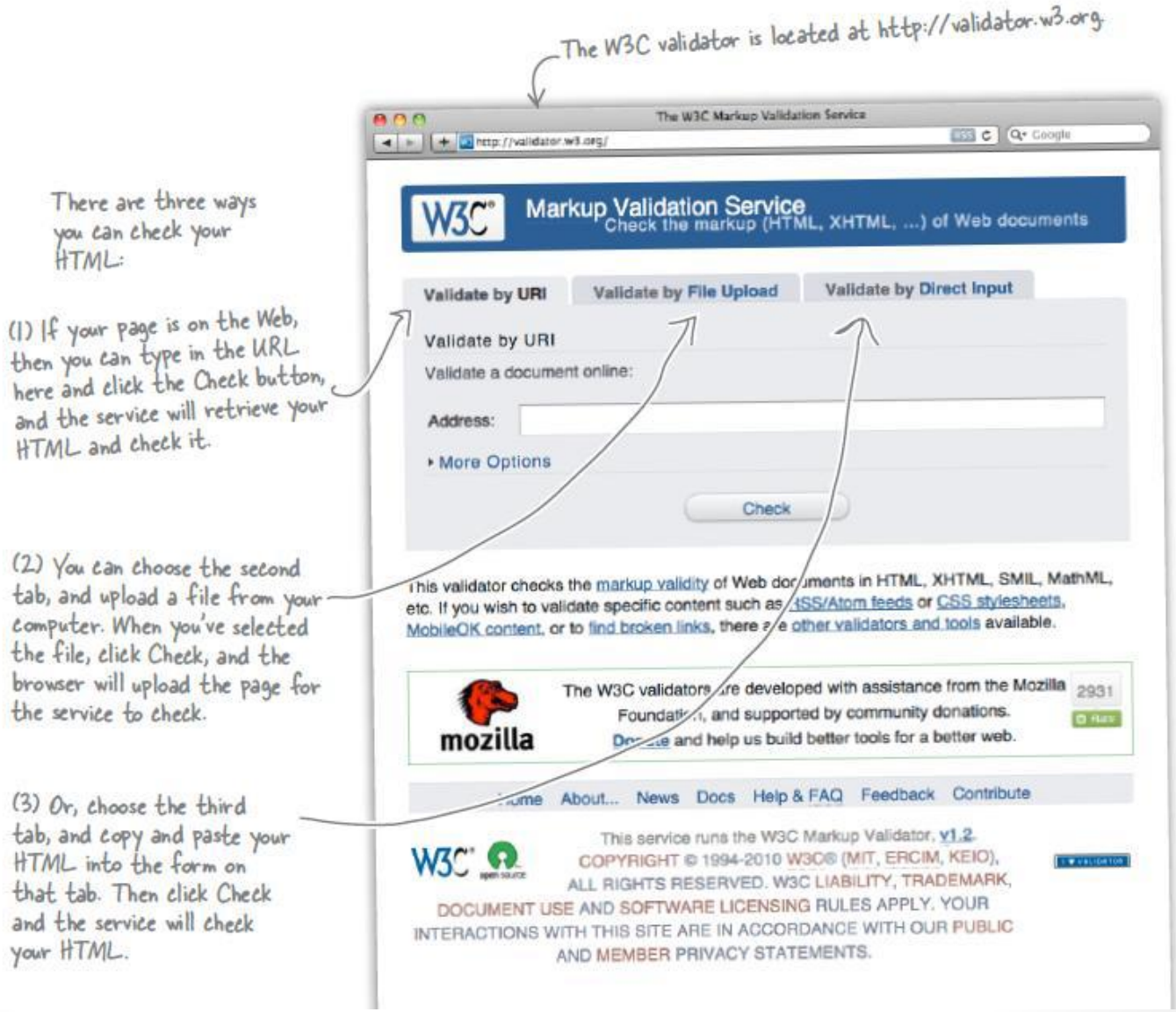


Wow, no difference. Well, we didn't really expect any because all the doctype does is let the browser know for sure you're using HTML5.

validador del *W3C*

- Es un servicio gratuito que ofrece el W3C
- Nos ayuda a comprobar la validez de los documentos web.
  - ***http://validator.w3.org/***

validador del *W3C*

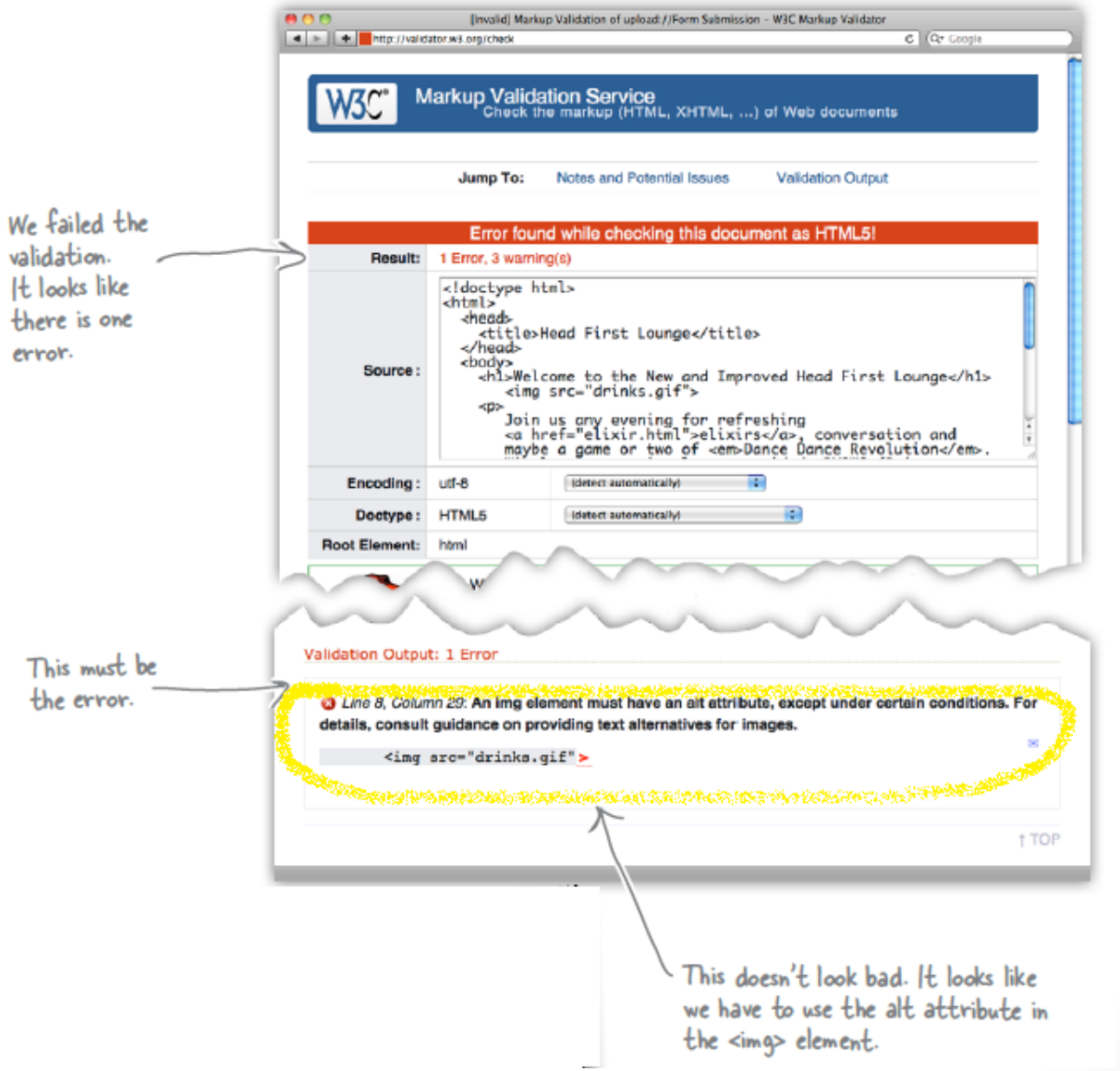


Validar lounge.html

We're using method (3) here. We clicked on the "Validate by Direct Input" tab and pasted the code for "lounge.html", which now has the doctype for HTML5 at the top, into the form. We're ready for the big moment...will the web page validate? Bets anyone? Click Check (and turn the page) to find out...

Feel free to use method (1) or (2) if it's more convenient.

Houston…tenemos un problema



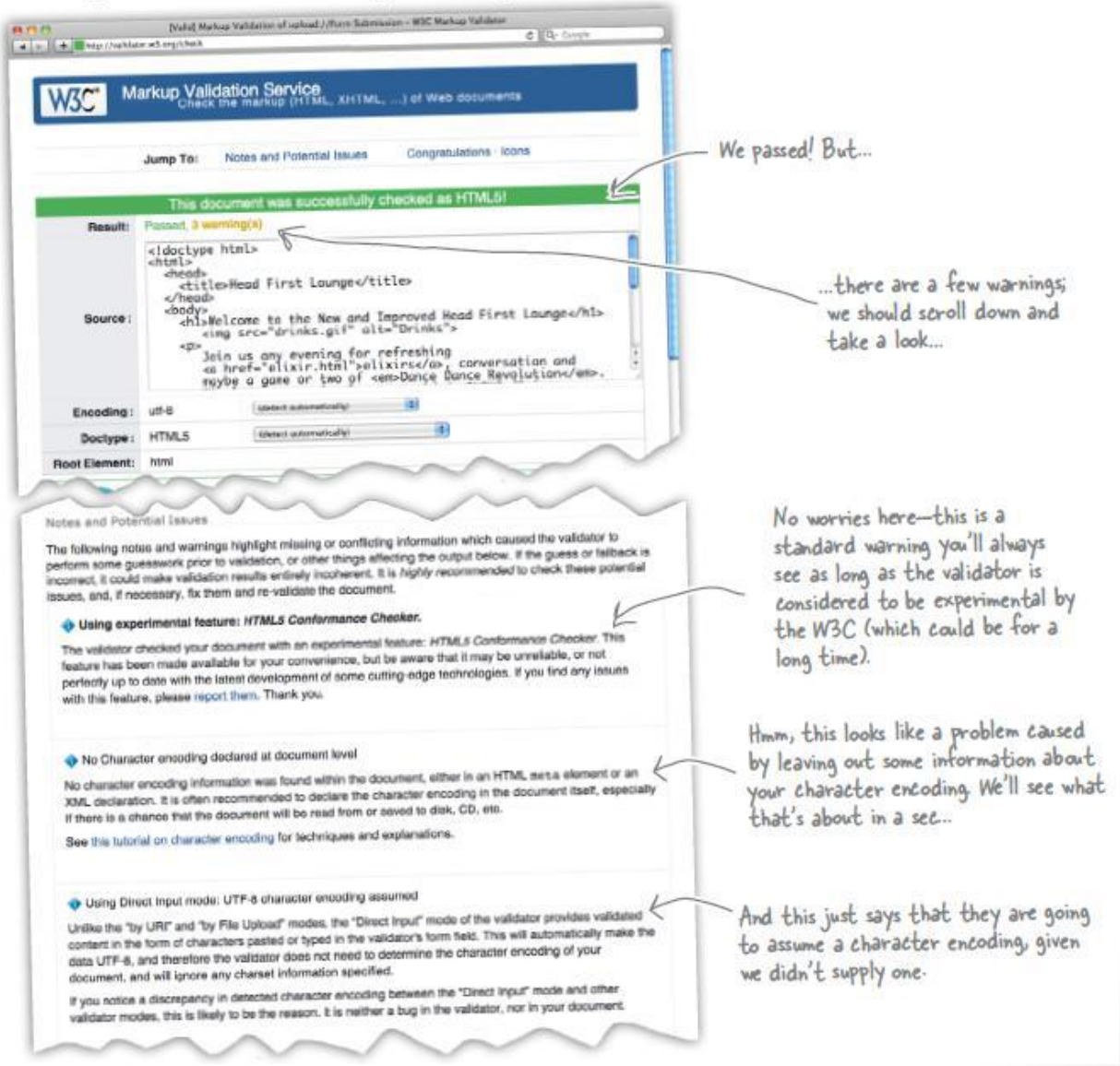
Corrigiendo el error

```
<!doctype html>
<html>
  <head>
    <title>Head First Lounge</title>
  </head>
  <body>
    <h1>Welcome to the New and Improved Head First Lounge</h1>
       <img src="drinks.gif" alt="Drinks">
    <p>
       Join us any evening for refreshing
       <a href="elixir.html">elixirs</a>, conversation and
       maybe a game or two of <em>Dance Dance Revolution</em>.
       Wireless access is always provided; BYOWS (Bring
       your own web server).
    </p>
    <h2>Directions</h2>
```

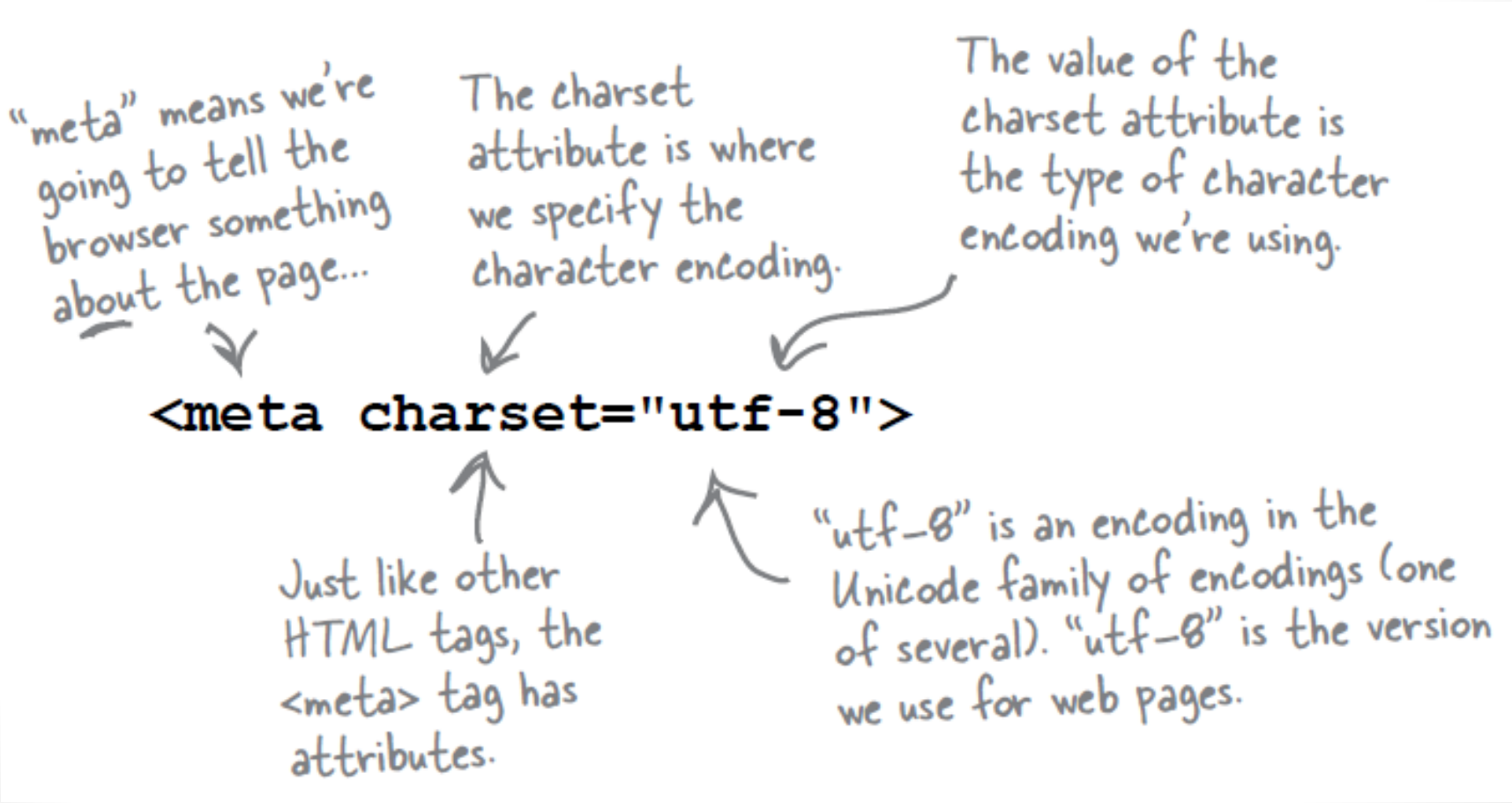
You know the alt attribute; add it into the <img> element.

Why do you think the alt attribute is required in HTML5?

Warnings


Markup Validation Service
Check the markup (HTML, XHTML, ...) of Web documents
Jump To: Notes and Potential Issues Congratulations Icons
This document was successfully checked as HTML5!
Result: Passed, 3 warning(s)
Source:
<!doctype html>
<html>
<head>
<title>Head First Lounge</title>
</head>
<body>
<h1>Welcome to the New and Improved Head First Lounge</h1>
<img src="drinks.gif" alt="Drinks">
<p>
Join us any evening for refreshing
elixirs, conversation and
maybe a game or two of Dance Dance Revolution.
Encoding: utf-8
Doctype: HTML5
Root Element: html
We passed! But...
...there are a few warnings; we should scroll down and take a look...
Notes and Potential Issues
The following notes and warnings highlight missing or conflicting information which caused the validator to perform some guesswork prior to validation, or other things affecting the output below. If the guess or fallback is incorrect, it could make validation results entirely incoherent. It is highly recommended to check these potential issues, and, if necessary, fix them and re-validate the document.
Using experimental feature: HTML5 Conformance Checker.
The validator checked your document with an experimental feature: HTML5 Conformance Checker. This feature has been made available for your convenience, but be aware that it may be unreliable, or not perfectly up to date with the latest development of some cutting-edge technologies. If you find any issues with this feature, please report them. Thank you.
No worries here—this is a standard warning you'll always see as long as the validator is considered to be experimental by the W3C (which could be for a long time).
No Character encoding declared at document level
No character encoding information was found within the document, either in an HTML meta element or an XML declaration. It is often recommended to declare the character encoding in the document itself, especially if there is a chance that the document will be read from or saved to disk, CD, etc.
See this tutorial on character encoding for techniques and explanations.
Hmm, this looks like a problem caused by leaving out some information about your character encoding. We'll see what that's about in a sec...
Using Direct Input mode: UTF-8 character encoding assumed
Unlike the "by URI" and "by File Upload" modes, the "Direct Input" mode of the validator provides validated content in the form of characters pasted or typed in the validator's form field. This will automatically make the data UTF-8, and therefore the validator does not need to determine the character encoding of your document, and will ignore any charset information specified.
If you notice a discrepancy in detected character encoding between the "Direct Input" mode and other validator modes, this is likely to be the reason. It is neither a bug in the validator, nor in your document.
And this just says that they are going to assume a character encoding, given we didn't supply one.

Etiqueta <meta>



etiqueta <meta>

Here's the <meta> tag. We've added it to the <head> element above the <title> element

Add this line above any other elements in the <head> element.

```
<!doctype html>
<html>
  <head>
    <meta charset="utf-8">
    <title>Head First Lounge</title>
  </head>
  <body>
    <h1>Welcome to the New and Improved Head First Lounge</h1>
      <img src="drinks.gif" alt="Drinks">
    <p>
      Join us any evening for refreshing
      <a href="elixir.html">elixirs</a>, conversation and
      maybe a game or two of <em>Dance Dance Revolution</em>.
      Wireless access is always provided; BYOWS (Bring
      your own web server).
```

Ejercicio

- Añade la etiqueta <meta> y el tipo de documento a todos los html del ejercicio lounge.
- Después pasa el validador y corrige todos los errores y warnings que aparezcan.

No hay preguntas tontas

there are no Dumb Questions

Q: Doctypes, <meta> tags...ugh, do I need to really do all this to write web pages?

A: Specifying a doctype and character encoding with a <meta> tag are kind of like taxes: you gotta do them to be compliant. Look at it this way: you already understand them more than 98% of the web page writing population, which is great. But at the end of the day, everyone just puts the doctype and <meta> tag in their HTML and moves on with life. So make sure you've got them in your HTML, and then go do something much more fun.

Q: utf-8?

A: Work with us here. It's like WD-40; you don't worry about why it's called that, you just use it. As we said, utf-8 (also written sometimes as UTF-8) is part of the Unicode encoding family. The *u* in utf-8 means Unicode. Unicode is a character set supported across many commonly used software applications and operating systems, and is the encoding of choice for the Web, because it supports all languages, and multilingual documents (documents that use more than one language). It's also compatible with ASCII, which was a common encoding for English-only documents. If you're interested in learning more about Unicode or character encodings in general, check out the information on character encoding at http://www.w3.org/International/O-charset.html.

Q: I've also seen <meta> tags that look like this: <meta http-equiv="Content-Type" content="text/html;charset=utf-8" >. Do I need to use this instead sometimes?

A: No. That is the format for the <meta> tag in HTML 4.01 and earlier. In HTML5, you can just write <meta charset="utf-8">.

Q: Is this why you had us save our files using utf-8 for the encoding way back in Chapter 1?

A: Yes. You want the encoding of the file you're serving to the browser to match the encoding you specify in the <meta> tag.

Buenas prácticas

Webville Guide to HTML

In this handy guide, we've boiled down writing well-formed HTML pages into a common sense set of guidelines. Check them out:

Always begin with the <doctype>.

Always start each page with a doctype. This will get you off on the right foot with browsers, and with the validator too.

Use `<!docytype html>` at all times, unless you really are writing HTML 4.01 or XHTML.

The `<html>` element: don't leave home without it.

Following the doctype, the `<html>` element must always be the top, or root, element of your web page. So, after the doctype, the `<html>` tag will start your page and the `</html>` tag should end it, with everything else in your page nested inside.

Buenas prácticas

Remember to use both your `<head>` and your `<body>` for better HTML.

Only the `<head>` and `<body>` elements can go directly inside your `<html>` element. This means that every other element must go either inside the `<head>` or the `<body>` element. No exceptions!

Feed your `<head>` the right character encoding.

Include a `<meta charset="utf-8">` tag in your `<head>`. The browser will thank you, and so will your users when they're reading comments on your blog from users around the world.

Buenas prácticas

Webville Guide to HTML, continued

In this handy guide, we've boiled down writing well-formed HTML pages into a common sense set of guidelines. Check them out:

What's a `<head>` without a `<title>`?

Always give your **`<head>`** element a **`<title>`** element. It's the law. Failure to do so will result in HTML that isn't compliant. The **`<head>`** element is the only place you should put your **`<title>`**, **`<meta>`**, and **`<style>`** elements.

Be careful about nesting certain elements.

Within the guidelines we've provided here, the nesting rules are fairly flexible. But there are a couple of cases that don't make sense. Never nest an **`<a>`** element inside another **`<a>`** element because that would be too confusing for our visitors. Also, void elements like **`<img>`** provide no way to nest other inline elements within them.

Buenas prácticas

Check your attributes!

Some element attributes are required, and some are optional. For instance, the `<img>` element wouldn't make much sense without a src attribute, and now you know the alt attribute is required too. Get familiar with the required and optional attributes of each element as you learn it.

Malas prácticas

```
<html>
<head>
   <title>Webville Forecast</title>
</head>

<body bgcolor="tan" text="black">


   <p>
      The weather report says lots of rain and wind in store for
      <font face="arial">Webville</font> today, so be sure to
      stay inside if you can.
   </p>

   <ul>
      <li>Tuesday: Rain and 60 degrees.
      <li>Wednesday: Rain and 62 degrees.
   </ul>

   <p align=right>
     Bring your umbrella!

   <center><font size="small">This page brought to you buy Lou's
       Diner, a Webville institution for over 50 years.
   </font></center>

</body>
</html>
```

Here are some attributes that controlled presentation. bgcolor sets the background color of the page, and text sets the color of the body text

Font changes were made with the <font> element and its face attribute.

You could get away without some closing tags, like and </p>. You sometimes still can, but it is not recommended!!

Missing quotes around attribute values. Quotes are always recommended now, and required for attributes with multiple values.

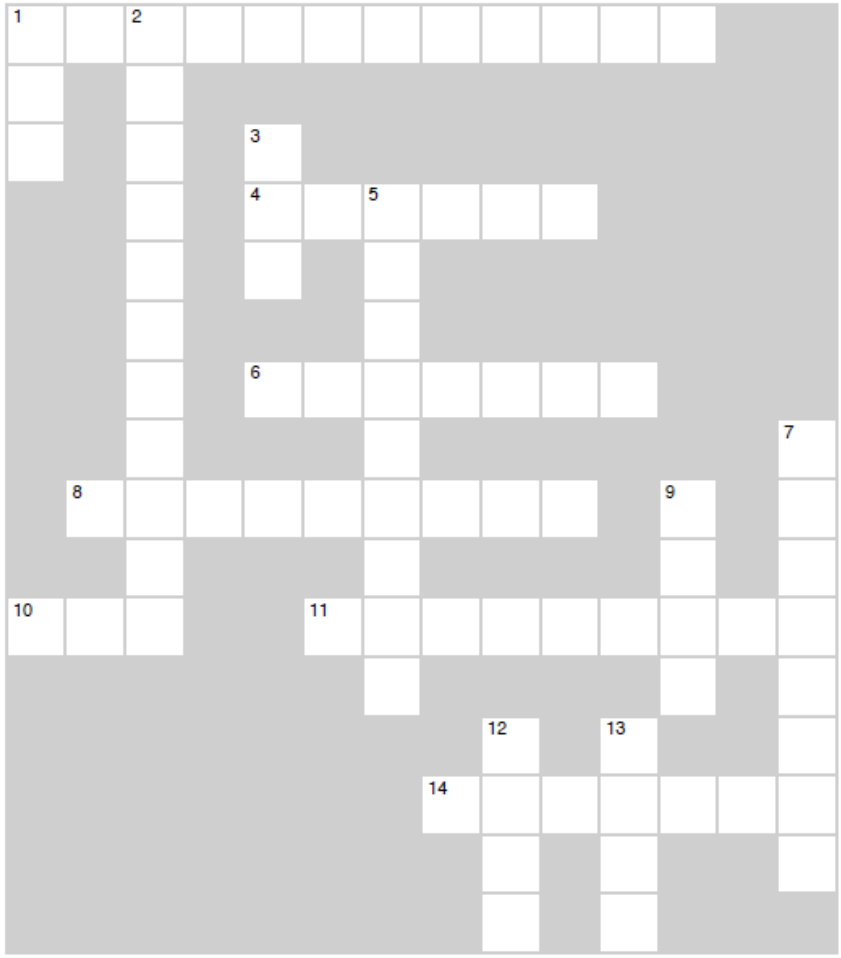
Here were two ways to align text. Right-align a paragraph, or center a piece of text

Text size was controlled with the <font> element, using the size attribute

Ejercicio: encuentra los errores

```
<html>
<head>
   <meta charset="utf-9">
</head>
<body>
   <img src="chamberofcommerce.gif">
   <h1>Tips for Enjoying Your Visit in Webville
   <p>
      Here are a few tips to help you better enjoy your stay in Webville.
   </p>
   <ul>
       <li>Always dress in layers and keep an html around your
           head and body.</li>
       <li>Get plenty of rest while you're here, sleep helps all
           those rules sink in.</li>
       <li>Don't miss the work of our local artists right downtown
           in the CSS gallery.
   </ul>
   </p>
   <p>
      Having problems? You can always find answers at
      <a href="http://wickedlysmart.com"><em>WickedlySmart</em></a>.
      Still got problems? Relax, Webville's a friendly place, just ask someone
      for help. And, as a local used to say:
   </p>
   <em><p>
      Don't worry. As long as you hit that wire with the connecting hook
      at precisely 88mph the instant the lightning strikes the tower…
      everything will be fine.
   </em></p>
</body>
</html>
```

Ejercicio



Across

1. Victim of the browser wars.
4. The HTML standard is a _____ standard.
6. Required in the <head> element.
8. Web standards makers have promised future HTML will be _____ compatible with older HTML.
10. The boss wanted to standardize before adding _____ to the Lounge pages.
11. When your HTML meets the standard, it is this.
14. Definition that tells the browser and validator what kind of document you're creating.

Down

1. Standards organization that supplies the validator.
2. Microsoft versus Netscape.
3. <img> attribute required in standard HTML.
5. This service will check your HTML for compliance with the standard.
7. The older _____ were much more complicated compared to the newest one.
9. Where you put information about the page.
12. Where you put web page content.
13. The most common encoding for web pages.

Puntos importantes

- HTML5 is the current HTML standard.
- The World Wide Web Consortium (W3C) is the standards organization that defines what standard HTML is.
- The document type definition (doctype) is used to tell the browser the version of HTML you're using.
- The HTML standard is now a "living standard," which means that the standard will change to incorporate new features and updates.
- The <meta> tag in the <head> element tells the browser additional information about a web page, such as the content type and character encoding.
- The charset attribute of the <meta> tag tells the browser the character encoding that is used for the web page.
- Most web pages use the utf-8 encoding for HTML files, and for the <meta> tag charset attribute.
- The alt attribute is required for the <img> element.
- The W3C validator is a free online service that checks pages for compliance with standards.
- Use the validator to ensure that your HTML is well formed and that your elements and attributes meet the standard.
- By adhering to the standard, your pages will display more quickly and with fewer display differences between browsers, and your CSS will work better.