**HTML5 articles and sections: what’s the difference?**

Thursday 15 April 2010

An article is an independent, stand-alone piece of discrete content. Think of a blogpost, or a news item.

Consider this real-world article:

<article>  
<h1>Bruce Lawson is World's Sexiest Man</h1>  
<p>Legions of lovely ladies voted luscious lothario Lawson as the World's Sexiest Man today.</p>  
<h2>Second-sexiest man concedes defeat</h2>  
<p>Remington Sharp, jQuery glamourpuss and Brighton roister-doister, was gracious in defeat. "It's cool being the second sexiest man when number one is Awesome Lawson" he said, from his swimming pool-sized jacuzzi full of supermodels.</p>  
</article>

It could be syndicated, either by RSS or other means, and makes sense without further contextualisation. Just as you can syndicate partial feeds, a “teaser” article is still an article:

<article>  
<a href=full-story.html>  
<h1>Bruce Lawson is World's Sexiest Man</h1>  
<p>Legions of lovely ladies voted luscious lothario Lawson as the World's Sexiest Man today.</p>  
<p>Read more</p>  
</a>  
</article>

Other articles can be nested inside an article, for example a transcript to a video:

<article>  
<h1>Stars celebrate Bruce Lawson</h1>  
<video>…</video>

<article class=transcript>  
<h1>Transcript</h1>  
<p>Priyanka Chopra: "He's so hunky!"</p>  
<p>Konnie Huq: "He's a snogtabulous bundle of gorgeous manhood! And I saw him first, Piggy Chops!"</p>  
</article>

</article>

The transcript is complete in itself, even though it’s related to the video in the outer article. The spec says “When article elements are nested, the inner article elements represent articles that are in principle related to the contents of the outer article.”

**SECTION**

Section, on the other hand, isn’t “a self-contained composition in a document, page, application, or site and that is intended to be independently distributable or reusable”. It’s either a way of sectioning a page into different subject areas, or sectioning an article into … well, sections.

Consider this article:  
  
<article>  
<h1>Important legal stuff</h1>  
<h2>Carrots</h2>  
<p>Thingie thingie lah lah</p>  
<h2>Parsnips</h2>  
<p>Thingie thingie lah lah</p>  
<h2>A turnip for the books</h2>  
<p>Thingie thingie lah lah</p>  
<strong>Vital caveat about the information above!</strong>  
</article>

Does the “vital caveat about the information above” refer to the whole article, eg everything under the introuctoryh1, or does it refer only to the information under the preceding h2 (“A turnip for the books”)? In HTML4, there is no way to tell. In HTML5, the section element makes its meaning unambiguous (and therefore, more “semantic”):

<article>  
<h1>Important legal stuff</h1>

<section>  
<h2>Carrots</h2>  
<p>Thingie thingie lah lah</p>  
</section>

<section>  
<h2>Parsnips</h2>  
<p>Thingie thingie lah lah</p>  
</section>

<section>  
<h2>A turnip for the books</h2>  
<p>Thingie thingie lah lah</p>  
</section>

<strong>Vital caveat about the information above!</strong>  
</article>

Now we can see that the vital caveat refers to the whole article. If it had been inside the final section element, it would unambiguously refer to that section alone. It would not have been correct to divide up this article with nested article elements, as they would not be independent discrete entities, which is why we used the section element.

OK. So we’ve seen that we can have article inside article and section inside article. But we can also have article **inside** section. What’s that all about then?

article inside section

Imagine that your content area is divided into two units, one for articles about llamas, the other for articles about root vegetables. (Or see today’s [**Guardian home page**](http://www.guardian.co.uk/) with its main news, a section of election picks, a section of “latest multimedia” etc).

You’re not obliged to markup your llama articles separately from your root vegetable articles, but you want to demonstrate that the two groups are thematically distinct, and perhaps you want them in separate columns, or you’ll use CSS and JavaScript to make a tabbed interface. In HTML4, you’d use our good but meaningless friend div. InHTML5, you use section which, like article invokes the HTML5 outlining algorithm, while div doesn’t because it has no meaning. (A great read on the outlining algorithm is Lachlan Hunt’s [**A Preview of HTML 5**](http://www.alistapart.com/articles/previewofhtml5)):

<div role=main>

<section>  
<h1>Articles about llamas</h1>

<article>  
<h2>The daily llama: buddhism and South American camelids</h2>  
<p>blah blah</p>  
</article>

<article>  
<h2>Shh! Do not alarm a llama</h2>  
<p>blah blah</p>  
</article>

</section>

<section>  
<h1>Articles about root vegetables</h1>

<article>  
<h2>Carrots: the orange miracle</h2>  
<p>blah blah</p>  
</article>

<article>  
<h2>Swedes: don't eat people, eat root vegetables</h2>  
<p>blah blah</p>  
</article>

</section>

</div>

Why not article? Because, in this example, each section is a collection of independent entities, each of which could be syndicated—but you wouldn’t syndicate the collection as an individual entity.

Note that a section doesn’t need to be lots of articles; it could be a collection of paragraphs explaining your creative commons licensing, an author bio or a copyright notice. In our example, each article could contain sub-articles or section, as explained above—or both.

**FINALLY, A CONCLUSION!**

[**Jeremy Keith writes that authors are confused about when to use the two elements**](http://adactio.com/journal/1654/). I think the namearticle is a cause of confusion; perhaps post or entry or even story would be more intuitive if you’re thinking about blog or news sites (although not all sites are like that, of course).

But I disagree that the two elements are so similar that they should be amalgamated. Jeremy writes

the only thing that distinguishes the definition of article from the definition of section is the presence of the phrase “self-contained”. A section groups together thematically-related content. An article groups together self-contained thematically-related content. That distinction is too fine to warrant a separate element, in my opinion.

I agree that the difference between them is the “self-contained”ness. But, personally, I find it pretty easy to work out whether something is self-contained or not and have tried to explain it above. Your comments will hopefully let me know if I’ve explained it clearly enough. (I think it’s very tough explaining it in the terse language required in normative sections of a specification).

It seems to me that brand-new elements will require people to spend time learning them without being able to immediately understand the difference in a matching exercise. Dan Cederholm’s [**Simplequiz**](http://simplebits.com/categories/simplequiz/) showed that in 2003 many of us didn’t understand HTML4 elements properly. How many of us would have chosen ol rather than ul from name and single line from the spec if asked the most appropriate element for breadcrumb trails, or chosen dt as the most appropriate term for the speaker’s name in a dialogue (as the HTML4 spec [**wrongly specifies**](http://www.brucelawson.co.uk/2006/breaking-news-w3c-specs-are-not-word-of-god/))? But seven years down the line, I imagine we all agree that it would have been wrong to amalgamate dl, ul and ol.

I also think the spec isn’t sufficiently clear (and [**emailed the Working Group**](http://lists.whatwg.org/htdig.cgi/whatwg-whatwg.org/2010-April/025928.html)): the definition for article says “The article element represents a self-contained composition in a document, page, application, or site and that is intended to be independently distributable or reusable, e.g. in syndication.”

This suggests that if you have a self-contained composition that you do not intend to be distributable via syndication, you shouldn’t use article.

Section says “Authors are encouraged to use the article element instead of the section element when it would make sense to syndicate the contents of the element” – here, the intent of syndication is diluted into “it would make sense to syndicate the content”.

I suggest that article be amended to say something similar, eg “The article element represents a self-contained composition in a document, page, application, or site which would make sense if independently distributed or reused, e.g. in syndication.” so that the two mentions of article match.

If we didn’t have an article element, we’d be left with lots of different riffs on section class=article,section class=story or section class=post, which is what HTML5 tries to avoid.